Available online at www.sciencedirect.com
SCIENCE DIRECT®
Progress in Biophysics and Molecular Biology ▮ (▮▮▮▮) ▮▮▮–▮▮▮
Progress in Biophysics & Molecular Biology
www.elsevier.com/locate/pbiomolbio

Review

Proteins that bind high-mannose sugars of the HIV envelope

Istvan Botos, Alexander Wlodawer*

Macromolecular Crystallography Laboratory, National Cancer Institute, NCI-Frederick, Building 536, Room 5, Frederick, MD 21702-1201, USA

Abstract

A broad range of proteins bind high-mannose carbohydrates found on the surface of the envelope protein gp120 of the human immunodeficiency virus and thus interfere with the viral life cycle, providing a potential new way of controlling HIV infection. These proteins interact with the carbohydrate moieties in different ways. A group of them interacts as typical C-type lectins via a Ca^{2+} ion. Another group interacts with specific single, terminal sugars, without the help of a metal cation. A third group is involved in more intimate interactions, with multiple carbohydrate rings and no metal ion. Finally, there is a group of lectins for which the interaction mode has not yet been elucidated. This review summarizes, principally from a structural point of view, the current state of knowledge about these high-mannose binding proteins and their mode of sugar binding.
Published by Elsevier Ltd.

Keywords: gp120; gp41; 2G12; Mannose-binding protein; Cyanovirin–N

Contents

*Corresponding author. Tel.: +1-301-846-5036; Fax: +1-301-846-6128.
E-mail address: wlodawer@ncifcrf.gov (A. Wlodawer).

0079-6107/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.pbiomolbio.2004.05.001

1. Introduction

According to the World Health Organization estimates, over 40 million people around the world were infected with human immunodeficiency virus (HIV) at the end of 2003, the latest year for which statistics became available as of the time of writing of this review (UNAIDS-WHO, 2003). Infection by HIV leads to the development of acquired immunodeficiency syndrome (AIDS), a disease that is usually slowly developing, but fatal if not treated. While initially found primarily among homosexuals and hemophiliacs, AIDS is a continuing threat to the general population—more than 75% of those infected acquired HIV through heterosexual contact. An estimated 5 million people worldwide were newly infected and 3 million died from AIDS-related complications in 2003 (UNAIDS-WHO, 2003). At this time, almost all of the anti-HIV drugs target reverse transcriptase or protease, two proteins encoded by the virus (Sansom and Wlodawer, 2000). The high mutation rate of primate immunodeficiency viruses is one of several challenges that researchers have encountered in the development of a human vaccine active against the multiple HIV clades (Burton, 1997). A new approach to antiretroviral therapy is to design inhibitors that target discrete steps in the viral entry pathway. Some of these inhibitors are already in clinical trials (Doranz et al., 2001; Kilby et al., 1998). An alternative route to AIDS prevention is the development of an anti-HIV virucide for topical or ex-vivo use (Lange et al., 1993).

To successfully develop anti-HIV virucides, the interactions between HIV and the cells of the human immune system must be clearly understood. Knowledge of the HIV envelope, the mechanisms of viral entry and viral propagation has grown enormously during recent years. However, there is no comprehensive review that focuses exclusively on the interactions between the carbohydrate moieties of the HIV envelope glycoproteins and proteins that are capable of binding these carbohydrates. Our goal is to combine the available structural information about proteins that bind the high-mannose sugars of the HIV envelope in order to promote a better understanding of the interactions of the key players, the envelope proteins and the carbohydrates, in the infection process.

2. The glycoproteins of HIV envelope

2.1. Viral infection and synthesis of HIV glycoproteins

The surface of HIV-1 virions is populated with glycoproteins (envelope or Env proteins). These proteins mediate the adsorption and penetration of host cells that are susceptible to infection. Specifically, HIV infection of target cells requires that the viral Env proteins bind sequentially to the immune receptor protein, CD4, and a coreceptor, CCR5 or CXCR4 (Berger et al., 1999; Doms and Peiper, 1997). These receptor-binding events trigger changes in the envelope glycoproteins that lead to membrane fusion and virus entry (Chan and Kim, 1998).

Retroviruses contain two Env glycoproteins, SU and TM, derived from a common precursor polypeptide. In the case of HIV-1, SU is often referred to as gp120, and TM is called gp41. In all HIV clades SU and TM are encoded by the *env* gene and are synthesized from a spliced version of the genomic RNA (see Fig. 1). The nascent Env polypetide binds to a signal recognition particle via its amino-terminal leader segment and then becomes associated with the membrane of the endoplasmic reticulum (ER). Further translation extrudes most of the polypeptide through the ER membrane into the ER lumen. The protein remains anchored in the ER membrane by a hydrophobic segment near the carboxyl terminus that spans the membrane once, with the carboxy-terminal tail of Env in the cytoplasmic compartment. In the ER, the Env protein forms a trimer (Coffin et al., 1997).

Concurrent with translation, the leader peptide sequence is removed and the protein becomes glycosylated. Modification of the N-linked oligosaccharides begins in the ER, as soon as they are added to the protein. All three of the glucose (Glc) residues and one of the nine mannose (Man) residues on each block of sugars are enzymatically removed. Before transit to the cell surface, the glycoproteins must be correctly folded and assembled into oligomers in the ER. To become fully fusogenic, Env proteins are cleaved in the Golgi by a cellular protease to yield mature SU and TM. N-linked oligosaccharides are extensively and often heterogeneously processed in the Golgi. Typically, three of the eight Man residues are removed, another GlcNAc residue is added, then another two Man residues are removed. In the trans-Golgi compartment many additional sugars (GlcNAc, Gal, sialic acid, Fuc) are added, further contributing to the heterogeneity of the resulting glycoproteins (Coffin et al., 1997).

After infection, the host cell synthesizes viral envelope glycoproteins encoded by the HIV-1 *env* gene as approximately 845–870 residue precursors in the rough ER (Earl et al., 1990). These

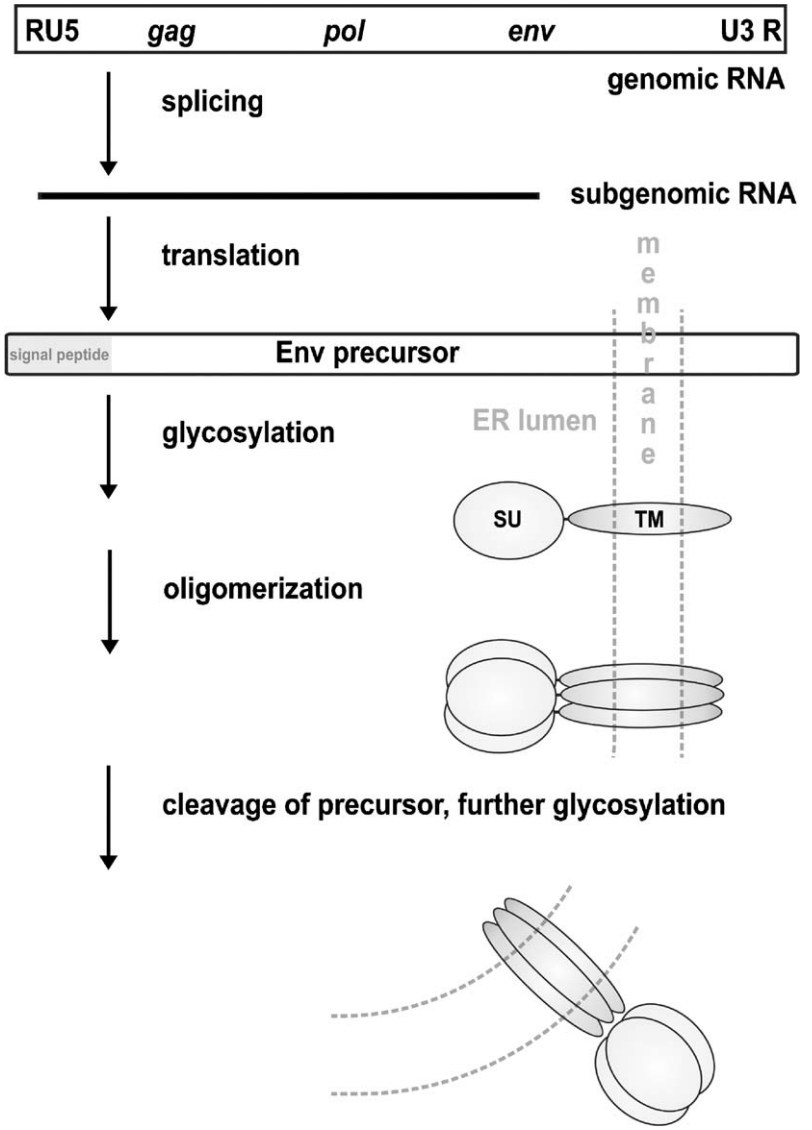

Fig. 1. Expression and processing of the HIV-1 envelope glycoproteins. This schematic diagram is adapted from Coffin et al. (1997).

precursors are *N*-glycosylated to form the gp160 glycoprotein by addition of identical, preassembled, $Glc_3Man_9GlcNAc_2$ blocks from a lipid intermediate (Earl et al., 1991; Leonard et al., 1990). Gp160 assembles into oligomeric complexes which are mainly trimers (Chan et al., 1997; Lu et al., 1995; Pinter et al., 1989; Weiss et al., 1990; Weissenhorn et al., 1997). These gp160 trimers are further processed in the Golgi apparatus by a cellular protease, and two mature envelope glycoproteins are generated—the external envelope glycoprotein, gp120, and the transmembrane glycoprotein, gp41 (Earl et al., 1990).

2.2. The external envelope glycoprotein gp120

The HIV-1 envelope glycoproteins gp120 and gp41 form a trimeric complex that mediates virus entry into target cells (Allan et al., 1985). A large fraction of the predicted accessible surface of gp120 in the trimer is composed of variable, heavily glycosylated core and loop structures that surround the receptor-binding regions (Wyatt et al., 1998).

Approximately 50% of the 120 kDa molecular weight of gp120 is provided by carbohydrate, all of it N-linked. No O-linked oligosaccharides were identified. Leonard et al. (1990) have shown that all 24 N-linked glycosylation sites are utilized in gp120, and that 13 of the sites contain complex-type oligosaccharides while 11 contain hybrid and/or high-mannose structures. The N-linked glycosylation sites in the gp120 amino acid sequence are well conserved (see Fig. 2), whereas there is a large variation in the amino acid sequence of gp120 from different viral isolates. A majority of the differences are localized in five hypervariable regions in the form of sequence variation, insertions, and deletions (Modrow et al., 1987; Willey et al., 1986). The variable loop regions are anchored at their bases by disulfide bonds that are highly conserved between all viral isolates except of Z3, which has an additional pair of cysteine residues in the fourth hypervariable loop (Leonard et al., 1990). The crystal structure of gp120 reveals the more conserved core region and five variable loop regions (V1–V5) (see Figs. 3A and B) (Wyatt et al., 1998). Neutralizing antibodies recognize discrete, conserved epitopes in three regions of the gp120 glycoprotein (see Table 1).

The structural studies of Wyatt et al. (1998) support the existence of neutralizing and non-neutralizing faces of gp120 and the presence of an immunologically "silent" face of the glycoprotein (see Fig. 3C). This "silent" face, together with the variable loops, forms a large part

```
                      90        100       110       120              200       210       220       230
                    c |         |         |         |              c  |         |         |         |
B HXBc2     garsEVVLVNVTENFNMWKNDMVEQMHEDIISLWDQSLKPCVKLTPLCVgagSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCNN
B YU2       ....EVKLENVTENFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVgagSCNTSVITQACPKVSFEPIPIHYCAPAGFAILKCND
C UG268A2   ....EIVLENVTESFNMWKNDMVDQMHQDVISLWDQSLKPCVKLTPLCV...NCNTSAITQACPKVTLDPIPIHYCAPAGYAILKCNN
O ANT70     ....EYPLHNVTDDFNIWKNYMVEQMQEDIISLWDQSLKPCVQMTFLCV...NCNSTTITQACPKVSFEPIPIHYCAPAGYAIFKCNS

                    240       250       260       270       280       290       330       340
                  m |       m |         | m       |     c   |       m |   m     | m     m |
B HXBc2     KTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVNFTDNAKTIIVQLNTSVEINCTgagHCNISRAKWNNTLKQIAS
B YU2       KKFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEIVIRSENFTNNAKTIIVQLNESVVINCT...HCNLSKTQWENTLEQIAI
C UG268A2   KTFNGTGPCNNVSTVQCTHGIKPVISTQLLLNGSIAEEEIIIRSENLTNNAKIIIVQLNKSVEINCA...YCNISRNEWNITLQWVRE
O ANT70     TEFNGTGTCRNITVVTCTHGIRPTVSTQLILNGTLSKGKIRMMAKDILEGGKNIIVTLNSTLNMTCE...YCNYNATDWGKILKQTAE

                    350       360       370       380       390       400       410       420       430
                      |     c   |         |         |     m   | m   c   |     c   |         |         |
B HXBc2     KLREQ..FGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTWSTEGSNNTEGSDTITLPCRIKQIINMWQKVGKA
B YU2       KLKEQ..FGNNKTIIFNPSSGGDPEIVTHSFNCGGEFFYCNSTQLFTWNDTRKLNNTGRNNNTEGSDTITLPCRIEQIINMWQKVG..
C UG268A2   KLKRH...FPNKTINFTQPSGGDLEITTHSFNCRGEFFYCNTSSLFNSSDNNNS...........TIITLPCRIKQIINMWQGVGRA
O ANT70     RYLELVNNTGSINMTFNHSSGGDLEVTHLHFNCHGEFFYCNTAKMFNYTFSCNGTTCSV.SNVSQGNNGTLPCKLRQVVRSWIRGQSG

                    440       450       460       470       480       490
                      |       m |         |     c   |         |         |
B HXBc2     MYAPPISGQIRCSSNITGLLLTRDGGNS..NNESEIFRPGGGDMRDNWRSELYKYKVVKIE
B YU2       ......RGQIRCSSNITGLLLTRDGGKD..TNGTEIFRPGGGDMRDNWRSELYKYKVVKIE
C UG268A2   MYAPPIKGKITCRSNITGLLLTRDGGETSETNSTETFRPGGGDMRDNWRSELYKYKVVEVK
O ANT70     LYAPPIKGNLTCMSNITGMILQMDNTW.NSSNNNVTFRPIGGDMKDIWRTELFNYKVVRVK
```

Fig. 2. Structure-based sequence alignment of HIV-1 gp120 variants showing the glycosylation sites. B: clade B, strain HCBc2; Y: clade B, strain YU2; C: clade C, strain UG268A2; O: clade O, strain ANT70. The glycosylation sites are marked c (complex) and m (high-mannose). Adapted from Kwong et al. (1998, 2000).

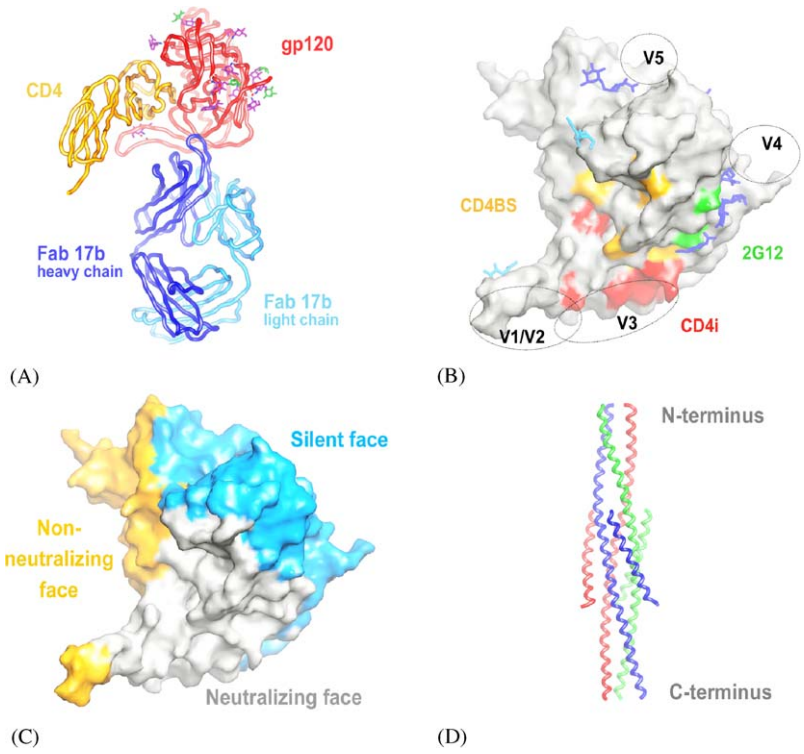

Fig. 3. (A) Ribbon diagram of gp120 bound to CD4 and Fab 17b (pdb code 1gc1, Kwong et al. (1998)). Gp120 is shown in red, CD4 in yellow and the light and heavy chains of Fab 17b antibody in light blue and dark blue, respectively. Asparagine residue-anchored surface sugars are shown in stick representation, *N*-acetyl glucosamine (NAG) in magenta, and fucose (Fuc) in green. Gp41 and the viral membrane are located towards the top of the diagram. Unless otherwise noted, color figures have been generated with program SPOCK (Christopher, 1998) (B) Molecular surface of HIV-1 gp120 core (pdb code 1gc1, Kwong et al. (1998)). Conserved gp120 neutralization epitopes for CD4 (yellow), CD4i (red) and 2G12 (green) are shown. The carbohydrates are shown in blue stick representation, with light blue denoting complex sugars and dark blue marking high-mannose sugars. The positions of the major variable loops (V1–V5) are shown. Adapted from Wyatt et al. (1998). (C) The approximate locations of the non-neutralizing (orange), neutralizing (gray) and the immunologically "silent" face (light blue) of gp120 (1gc1). The silent face roughly corresponds to the highly glycosylated outer domain surface. Adapted from Wyatt et al. (1998). (D) Ribbon diagram of the gp41 trimeric core (pdb code 1fav, Zhou et al. (2000)). Each monomer consists of: 29 residues of GCN4 fused to residues 30–77 of gp41 and residues 117–154 of the outer-layer helix of gp41.

Table 1
gp120 core neutralizing antibodies (based on Wyatt et al., 1998)

Epitopes	Monoclonal antibodies	Sugars involved	References
CD4BS	F105, 15e, 21h, 1125h, 448D, 38.3, IgG1b12, 830D	Possibly	Ho et al. (1991), Posner et al. (1991) and Thali et al. (1992)
CD4i	17b, 48d	Some complex sugars	Thali et al. (1993)
2G12	2G12	High-mannose	Trkola et al. (1995)

of the gp120 surface that is protected against antibody responses by rapid mutations in the variable loops and the dense array of carbohydrates that blocks antibody access to conserved regions of the gp120 core (Wyatt et al., 1998).

2.3. The transmembrane glycoprotein gp41

Gp41 consists of a 172 residue ectodomain, a 21 residue membrane-spanning anchor and a 142 residue cytoplasmic tail. The amino-terminal 129 residues of gp41 are required for efficient oligomer formation and stability (Earl and Moss, 1993), with residues 33–82 forming a leucine zipper-like domain (Shugars et al., 1996). The trimer of the gp41–gp120 complex is maintained by somewhat labile, non-covalent interactions between the highly conserved ectodomain of gp41 and amino- and carboxy-terminal sequences of gp120 (see Fig. 3D). These two terminal regions in gp120 are the most highly conserved between primate immunodeficiency viruses (Helseth et al., 1991).

Non-neutralizing antibodies of gp41 are directed against the immunodominant region following the leucine zipper-like domain. Monoclonal antibodies directed against other regions are rare and have weak neutralizing capacities (Nixon et al., 1992). In contrast with gp120, there are no antibodies identified that are directed against the surface glycoproteins of gp41.

3. Oligosaccharides of the HIV-1 envelope

The structures of the basic carbohydrate building blocks of oligosaccharides and some of the branched oligosaccharides mentioned in this review are summarized in Fig. 4. The axial and equatorial positions for substituents in a pyranose ring are also shown.

In the most common asparagine-linked oligosaccharides, the reducing end of the pentasaccharide $Man_3GlcNAc_2$ is linked to the side chain amide nitrogen of the asparagine residue. This is called the core pentasaccharide, and the three terminal mannose residues are referred to as trimannosidic core (Rao et al., 1998).

N-linked carbohydrates that share a common pentasaccharide core can be classified into three categories: high mannose, hybrid, and complex (Kornfeld and Kornfeld, 1985). All three types of oligosaccharides can be found in the gp120 protein of the HIV-1 envelope. The structure of these oligosaccharide moieties has been extensively analyzed; 33% were found to be of the high-mannose type, 4% of the hybrid type, and 63% of the complex type (Mizuochi et al., 1988a, b). Of the complex oligosaccharides, 90% were fucosylated and 94% were sialylated. 4% of the complex structures were monoantennary, 61% biantennary, 19% triantennary and 16% tetraantennary. High-mannose oligosaccharides have two to six mannose residues attached to this core, whereas complex oligosaccharides have different numbers and types of residues in their outer branches. These branches can be bi-, tri- and tetra-antennary chains with a typical sialyllactosamine sequence, and may have a fucose residue attached to the inner *N*-acetylglucosamine residue (Sanders et al., 2002). Hybrid oligosaccharides contain elements of both high-mannose and complex carbohydrate structures in their $1\rightarrow6$ branch or $1\rightarrow3$ branch, respectively. The carbohydrate moieties are transferred to asparagine residues in the rough endoplasmic reticulum and then trimmed to yield $Man_8GlcNAc_2$ oligosaccharides. These hybrid oligosaccharides can be further processed in the Golgi apparatus to become complex-type carbohydrates (Kornfeld and Kornfeld, 1985). Complex

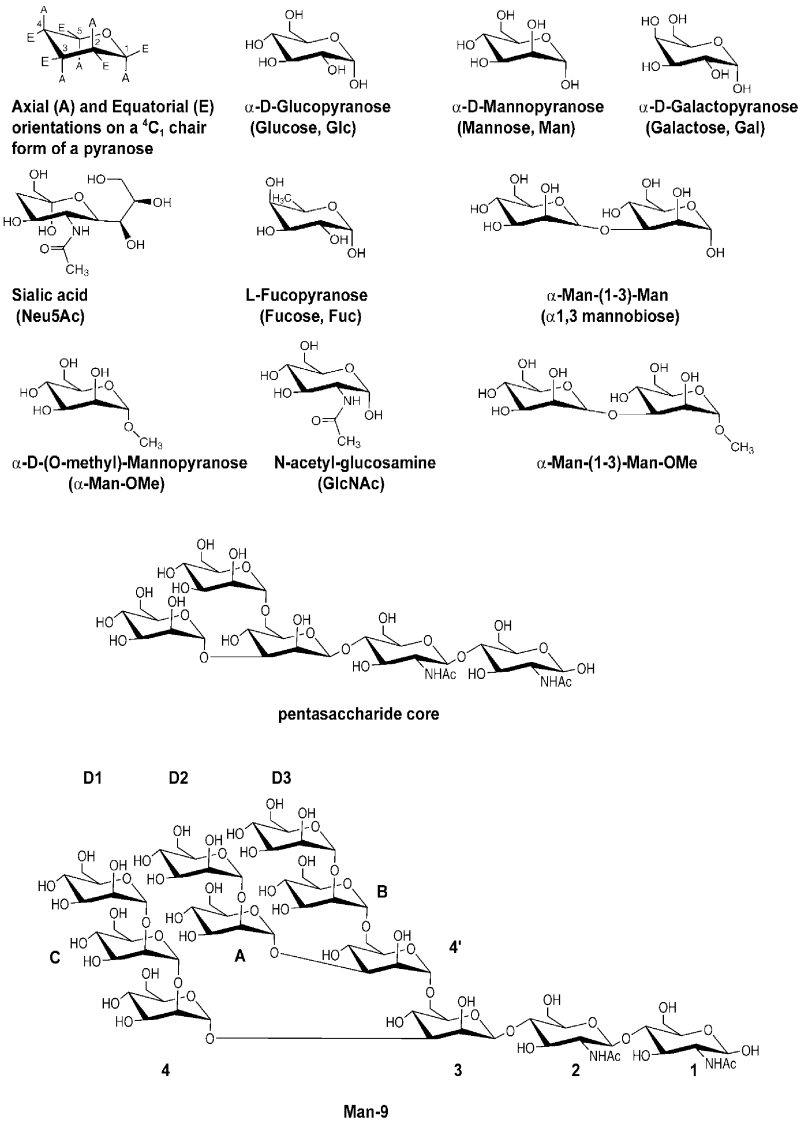

Fig. 4. Nomenclature, numbering, and chemical structure of some mono- and oligosaccharides.

oligosaccharides can contain Man, Gal, GlcNAc, GalNAc, L-Fuc and sialic acids in their branches. They occur as di-, tri-, tetra-, and penta-antennary sugar structures (Rao et al., 1998).

N-linked glycans, or sugars, differ from typical amino acid side chains in several respects. The average molecular weight of N-linked glycans is more than 20 times that of a typical amino acid side chain. They present greater structural complexity, greater bulk, and affect a larger volume of their immediate environment (Sanders et al., 2002). The total surface area of a single pentasaccharide is approximately the size of an antibody footprint.

The available crystal structures of oligosaccharides reveal only a limited number of conformations for glycosidic linkages (Petrescu et al., 1999). Molecular dynamics simulations

and NMR results suggest that branched high-mannose carbohydrates, despite their flexible glycosidic linkages, have well-defined overall conformations (Woods et al., 1998). The restricted oligosaccharide orientation might be functionally important in the interaction with sugar-binding lectins. For example, a second lectin–carbohydrate interaction may be enhanced by the reduced conformational space available and the reduced entropic penalty created by the first lectin–carbohydrate interaction. Similar mechanisms are likely to be involved in multivalent carbohydrate–lectin interactions (Ng et al., 2002).

In contrast to the glycoproteins of the HIV-1 envelope, mammalian cell surface or serum glycoproteins rarely contain terminal mannose residues (Weis et al., 1998). In mammals, the presence of a terminal mannose moiety on a protein triggers their binding to hepatic lectin receptors and is followed by rapid removal from the plasma (Lee et al., 2002). Virions expressing HIV-1 envelope glycoproteins have a half-life of minutes in macaque plasma (Igarashi et al., 1999). The presence of high-mannose terminal residues on gp120 is paradoxical. The protein has a highly conserved amino acid sequence, which suggests a precise, biologically relevant function, yet the presence of high mannose sugars, which trigger rapid removal of the protein, is detrimental to the propagation, and hence, the survival, of the virus. Ultimately, HIV-1 uses these terminal mannose residues to its advantage against the human immune system by reducing its antigenicity (Sanders et al., 2002).

4. High-mannose binding proteins of the HIV envelope

In this review we will discuss in detail only those proteins for which the mode of their interaction with HIV-1 glycoproteins has been proven to involve high-mannose sugars. However, some proteins that might interact with sugars upon glycoprotein binding will be briefly mentioned as well. The proteins that meet these criteria range from antibodies like 2G12 to a series of lectins. Depending on their origin, high-mannose binding lectins can be from plants (ConA, UDA, jacalin, GNA, SCL, NPA), animals (C-type lectins such as DC-SIGN or MBL), or from blue-green algae (cyanovirin–N, scytovirin).

The first of these proteins is the CD4 receptor that itself is a glycoprotein. CD4 interacts with gp120 via the CD4-binding site (CD4BS), which is devoid of carbohydrates (Wyatt et al., 1998) and formed by a set of conserved residues that have been identified by deletion mutagenesis (see Figs. 3A and B). The central motif of this well-conserved binding site is the "Phe43" cavity, a 10 Å pocket into which the critical Phe43 of CD4 inserts upon binding (Peterson and Seed, 1988; Wyatt et al., 1998). This recessed pocket is quite deep, even in the core gp120 structure, in which the large, flexible branches of the N-linked sugars were enzymatically removed. With the intact sugars present, one can envision a rather deep pocket where CD4 binds and possibly partially interacts with two of the complex and one of the high-mannose sugar branches on the rim of the binding cavity. However, enzymatic deglycosylation studies in the presence or absence of detergents showed no loss of CD4-binding affinity upon removal of most of the glycosyl moieties (Fenouillet et al., 1989; Lifson et al., 1986; Matthews et al., 1987).

Another group of proteins that might interact with high-mannose surface sugars upon gp120 binding are the CD4BS and CD4i antibodies (see Table 1). These are neutralizing monoclonal antibodies with specificity for a broad range of sugars. The chemokine receptor CCR5 also

interacts with gp120 via an epitope similar to that of CD4i antibodies (Trkola et al., 1996a). The exact mode of interaction is unknown, and long branches of N-linked high-mannose sugars might contribute to the binding of this receptor.

4.1. The 2G12 antibody

2G12 is one of the few broadly neutralizing monoclonal antibodies directed against HIV-1. This mAb recognizes an epitope around the C4/V4 region of gp120 with nanomolar affinity (Kunert et al., 1998; Trkola et al., 1996b). In vitro, 2G12 neutralizes a wide spectrum of different HIV-1 isolates, from all clades except clade E (Trkola et al., 1995, 1996b). In vivo, 2G12 protects macaques from the challenge by simian–human chimeric immunodeficiency virus (SHIV) 89.6P (Mascola et al., 2000).

The crystal structure of the gp120 core shows that this epitope is located on the relatively variable surface of the outer domain, opposite to the CD4-binding site (Wyatt et al., 1998) (see Fig. 3B). Mutations that alter attachment of N-linked carbohydrate chains to gp120 can disrupt the integrity of the 2G12 epitope (Trkola et al., 1996b). Even in the enzymatically deglycosylated gp120 core structure, this epitope is populated with N-linked *N*-acetylglucosamines. Mammalian cells densely glycosylate these regions, mainly with high-mannose sugars, and form an extended solvent-accessible surface expected to be weakly immunogenic, therefore designated as the "silent" face (Kwong et al., 1998, 2000; Wyatt et al., 1998; Wyatt and Sodroski, 1998). 2G12 likely binds to these high-mannose sugars. The epitope is highly conserved in spite of the variability of the underlying protein surface (Wyatt et al., 1998).

Carbohydrate-rich regions of glycoproteins are considered poor immunogens for a number of reasons. First, carbohydrate moieties exhibit microheterogeneity. The same protein sequence can exhibit a broad range of glycoforms, causing the dilution of any single antigenic response (Rudd and Dwek, 1997). Second, large carbohydrates are flexible and extend considerably from the protein core, being able to cover potential epitopes (Wilson et al., 1981). Third, viruses depend on the host glycosylation machinery, therefore the oligosaccharides attached to potential antigens are the same as those attached to host glycoproteins (Scanlan et al., 2002). One can assume that, in general, the host organism will better tolerate these carbohydrates. This difficulty in eliciting antibodies to a carbohydrate face is consistent with the special nature of 2G12.

2G12 forms a unique competition group (Moore and Sodroski, 1996; Trkola et al., 1996b) and does not interfere with the binding of monomeric gp120–CD4 or chemokine receptors (Trkola et al., 1996a); in fact, no other monoclonal antibody is able to prevent its binding to gp120, and vice versa. Upon CD4 binding, the 2G12 epitope is predicted to be oriented towards the target cell. Such an orientation might impair the interactions of the trimeric envelope glycoprotein complex with the host cell.

The nature of previously identified residues of the 2G12 epitope (Trkola et al., 1996b) was further investigated through sequence analysis of HIV-1 isolates (Sanders et al., 2002). The results of alanine-scanning mutagenesis and asparagine-to-glutamine substitutions in $gp120_{JR-FL}$ were mapped onto the crystal structure of the gp120 core (Scanlan et al., 2002). The residues supporting the N-linked glycans directly involved in 2G12 binding are Asn295, Asn332, Asn392, whereas Asn339 and Asn386 glycans influence 2G12 binding. Sequence alignment of gp120 proteins from HIV-1 isolates efficiently neutralized by 2G12 showed that the N-linked carbohydrate signal

sequences associated with Asn295, Asn332, and Asn392 are highly conserved (see Table 1) (Scanlan et al., 2002; Zhu et al., 2000).

The possible involvement of the V1–V4 loops of gp120 in 2G12 binding was investigated because of their close proximity to the carbohydrate moieties. The sequence variability of loop V4 makes it less likely to have a direct role in 2G12 binding. However, V4 main chain residues may be indirectly involved in 2G12 binding to gp120 by maintaining the relative positions of the interacting carbohydrate chains (Scanlan et al., 2002). The importance of the conservation of the secondary structure of the V4 loop for recognition of gp120 is also supported by the observation that clade E isolates, which have an additional disulfide bond in the V4 loop, are not recognized by 2G12 (Trkola et al., 1995). Loops V1–V3 had minimal effect on antibody binding (Scanlan et al., 2002).

Exoglycosidase and endoglycosidase digestion studies showed that removal of the mannose residues by endoH dramatically reduced the binding affinity of 2G12–gp120. Removal of α-Man-(1→2)-α-Man-linked residues by *A. saitoi* mannosidase (leaving $Man_5GlcNAc_2$ structures) or α-Man-(1→2,3,6)-α-Man-linked residues by jack bean mannosidase (leaving $Man_1GlcNAc_2$ structures) reduced the binding affinities of both 2G12 and cyanovirin to gp120 drastically, but not that of the anti-CD4BS antibody b12. The 2G12 epitope is either formed exclusively of the outer α-Man-(1→2)-α-Man residues of the glycosyl moieties or also involves α-Man-(1→2,3,6)-α-Man residues in the context of α-Man-(1→2)-α-Man residues. Consistent with these results, high concentrations of *D*-mannose inhibited the gp120–2G12 interaction, while glucose, galactose, and *N*-acetylglucosamine had little effect. The precise location of the C2 hydroxyl group in mannose might be a key determinant of 2G12 specificity, since mannose differs from glucose only in the stereochemistry of this hydroxyl group (Scanlan et al., 2002).

The glucose analogue *N*-butyldeoxynojirimycin (NB-DNJ) inhibits activity of the endoplasmic reticulum glucosidases I and II, enzymes which remove the terminal glucose residues attached to the D1 arm of the precursor $Glc_3Man_9GlcNAc_2$. Inhibition of these glucosidases leads to the secretion of mainly triglucosylated oligomannose-type oligosaccharides (Karlsson et al., 1993), and inhibition of any 2G12 binding to gp120. The additional glucose residues on the D1 arm might sterically hinder the 2G12 binding. This observation further emphasizes the importance of the terminal α-Man-(1→2)-α-Man residues in 2G12 binding (Scanlan et al., 2002). It has been shown previously that NB-DNJ can affect the calnexin-mediated folding of glycoproteins in the ER that can lead to misfolding of the V2 loop in gp120 (Fischer et al., 1996). The $gp120_{JR-FL}$ N-linked carbohydrates were also analyzed by normal-phase HPLC after enzymatic release and fluorescence labeling (Scanlan et al., 2002). The abundance of oligomannose species on $gp120_{JR-FL}$ (Table 2) was in agreement with a previous, mass-spectrometric study of oligomannose distribution of $gp120_{SF2}$ (Zhu et al., 2000).

The crystal structures of unliganded and saccharide-bound Fab 2G12 revealed a tightly packed dimer formed by V_H domain swapping (Calarese et al., 2003) (see Fig. 5A). Such V_H domain swapping is unique, since despite the structural similarity of domains V_H, V_L, C_H1, and C_L with other Fab molecules, it has not been reported in more than 250 Fab structures available to date. The domain exchange is achieved by twisting the variable regions with respect to the constant region, relative to their standard orientation in a Fab (Calarese et al., 2003). The two Fabs are arranged side-by-side with their respective combining sites facing in the same direction and separated by ~35 Å. Structural analysis of Fab 2G12 identified three key factors that might

Table 2
Relative abundance of gp120 glycosyl species on the isolate JR–FL (based on data from Scanlan et al. (2002))

Type of saccharide	Nomenclature	Relative abundance, %	Total
Neutral complex	A2G2F	6.6	10.84
	A3G3	1.9	
	A4G0	1.3	
	A4G0F	0.7	
	A4G4F	0.34	
Sialylated complex	A2G2FS	13.5	22.0
	A2G2S	4.0	
	A3G3S3	1.9	
	A3G3S2	1.6	
	A4G4S4	<1	
Oligomannose	Man-8	20.6	58.4
	Man-7	14.3	
	Man-6	10.8	
	Man-9	10.3	
	Man-5	2.4	

The abbreviations are: G: glucose, S: sialic acid, F: L-fucose, A: galactose. For the complex type oligosaccharides the monosaccharides in the abbreviation are attached to the common $Man_3GlcNAc_2$ core (omitted).

facilitate domain swapping (Liu and Eisenberg, 2002): weakening of the V_H/V_L interface contacts, a Pro residue in the hinge loop connecting V_H and C_H1, and the formation of a favorable $V_H/V_{H'}$ interface (Calarese et al., 2003). The V_H/V_L interface is perturbed by the absence of a highly conserved hydrogen bond between the V_H and V_L domains, caused by the presence of a rarely occurring Arg residue. The $Pro113_H$ residue in the hinge region is also unusual and is stabilized by hydrophobic interactions with a $Val84_H$ residue (H refers to the heavy chain of the antibody). The $V_H/V_{H'}$ interface is characterized by shape complementarity, an extensive hydrogen bonding and salt bridge network, and 136 van der Waals interactions. A molecular surface of 1245 $Å^2$ is buried at this interface (Calarese et al., 2003). The dimeric state of 2G12 in solution was confirmed by sedimentation equilibrium analytical ultracentrifugation and by gel filtration. Negative-stain electron microscopy of intact 2G12 IgG1, as well as 2G12 bound to a covalently constrained gp120/gp41 molecule (Schulke et al., 2002), showed an unusual linear conformation of the antibody, with a domain-swapped Fab.

The crystal structures of 2G12 bound to $Man_9GlcNAc_2$ (see Fig. 5A) and α-Man-(1→2)-Man show that the 2G12 dimer contains multiple, distinct carbohydrate binding sites: two for the normal antibody site and two novel sites within the new $V_H/V_{H'}$ interface. In the primary binding sites, 2G12 contacts four sugars (3, 4, C and D1) in the D1 arm of the $Man_9GlcNAc_2$ complex. The terminal α-Man-(1→2)-Man disaccharide accounts for the majority of these contacts. In the disaccharide complex, α-Man-(1→2)-Man occupies only the two conventional binding pockets, located ~35 Å apart. 2G12 interfaces with α-Man-(1→2)-Man through 12 hydrogen bonds and 48 van der Waals interactions in each sugar-binding pocket via residues 93_L–94_L, 31_H–33_H, $52A_H$, and 95_H–$100D_H$ (L and H denote the light and heavy chains, respectively) (see Fig. 5B). The disaccharide buries 226 $Å^2$ of the 2G12 molecular surface (Calarese et al., 2003).

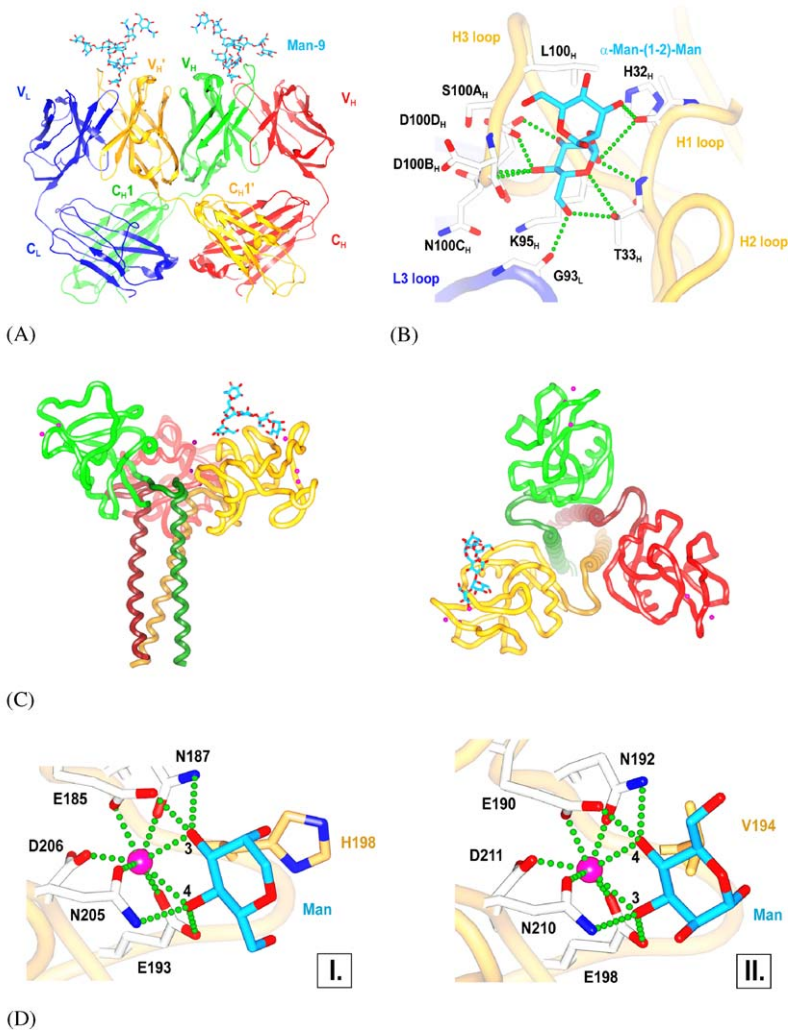

Fig. 5. (A) Ribbon diagram of two domain-swapped Fab 2G12 molecules with Man-9 bound to the two conventional pockets (pdb code 1op5, Calarese et al. (2003)). The monomers (blue + green and red + yellow) consist of a heavy (H) and a light (L) chain, with their variable heavy (V_H) domains swapped. The oligosaccharide is shown in stick representation (light blue). (B) Fab 2G12 complexed to α-Man-(1→2)-Man (1op3, Calarese et al., 2003). The carbohydrate (light blue) and the residues involved in its binding are shown in stick representation, with the hydrogen bonds shown as green spheres. Adapted from Calarese et al. (2003). (C) Side and top view ribbon diagrams of MBP–A complexed with $Man_6GlcNAc_2$ (pdb code 1kx1, Ng et al., 2002). The individual molecules of the trimer are color-coded. The Ca^{2+} atoms are shown as purple spheres and the bound carbohydrate $Man_6GlcNAc_2$ is shown in stick representation (light blue). (D) Orientations of carbohydrates bound to mannose-binding proteins. Orientation I: the terminal mannose of Manα1-3Man is shown complexed to serum MBP–A (pdb code 1kwy, Ng et al. (2002)). Orientation II: the same terminal mannose complexed to liver MBP–C (pdb code 1kza, Ng et al. (2002)). Ca^{2+} is depicted as a purple sphere, with coordinating residues in stick representation. His189 and Val194 (both in orange) participate in van der Waals interactions. The Ca^{2+} coordination and hydrogen bonds between protein and carbohydrate (blue) are shown as green spheres.

Binding-competition studies showed that interaction of D1 and C mannose with the primary sites of 2G12 cannot make up for the large affinity increase observed upon $Man_9GlcNAc_2$ binding. 2G12 $Asp100B_H$ and $Tyr94_L$ can form additional hydrogen bonds with branching mannose residues 3 and 4 from the D1 arm, respectively, and possibly increase the binding affinity. The surface area buried by bound $Man_9GlcNAc_2$ is also larger, 350–450 $Å^2$.

Several elements contribute to the specificity of the primary sugar-binding sites on 2G12. First, the deep pocket can accommodate only terminal sugar residues, and its specific shape is favorable for hydrogen bonding with α-Man-(1→2)-Man moieties. Second, additional interactions with mannose 3 and 4 from the D1 arm provide specificity for interaction with D2 arm terminal mannoses.

A secondary binding site is formed by the new $V_H/V_{H'}$ interface as a result of domain swapping. This site interacts mainly with the central mannose A of the D2 arm, but it also contacts the D2 and 4′ sugars. The binding site is relatively shallow and accommodates a carbohydrate chain of the D2 arm that runs parallel to the antibody surface. In solution, D1 or D3 arms could also interact with the secondary binding site. The identity of the 2G12 residues involved in these two sugar-binding pockets was also confirmed by alanine mutagenesis of the key amino acid residues. These mutation studies also emphasized the possible role of the secondary binding site in multivalent binding of 2G12–gp120 (Calarese et al., 2003).

The structure of 2G12 represents an elegant immunological solution to the recognition of carbohydrates on HIV-1. No known mammalian glycoprotein has such a dense cluster of carbohydrates as that observed on gp120. This characteristic distinguishes it from the cell's own glycoproteins. It is also distinct because a conventional Y- or T-shaped IgG molecule could not achieve a bivalent binding mode, due to the fact that a single sugar-binding site on the antibody can only interact with carbohydrates from a single oligomannose chain. In the near parallel, energetically unfavorable, orientation of the two Fab arms, an IgG molecule could bivalently bind two oligomannose chains 35 Å apart. In contrast, the domain-swapped 2G12 is well adapted for this task, and it provides nanomolar affinity recognition for a carbohydrate antigen. The available conformational space of $Man_9GlcNAc_2$–Asn, anchored by the *N*-glycosidic linkage, suggests that the Fab 2G12 dimer most likely interacts with gp120 at the N-linked glycans at positions Asn332 and Asn392 in the primary combining sites and possibly at position Asn339 at the $V_H/V_{H'}$ interface. Based on this model, the Asn295 glycan plays an indirect role in 2G12 binding, by preventing further processing of the glycan at Asn332 and maintaining its conformation (Calarese et al., 2003).

The binding of DC-SIGN to gp120 is competitive to that of 2G12 and possibly interferes with it through steric hindrance. In contrast to other C-type lectins, DC-SIGN binds internal trisaccharides of high-mannose oligosaccharides and does not bind terminal α-Man-(1→2)-Man residues (Feinberg et al., 2001). DC-SIGN binds a different ligand than 2G12, and its interaction with gp120 is perturbed by the presence of EDTA, unlike the interaction of 2G12 (Scanlan et al., 2002).

Cyanovirin–N (CV–N) can block 2G12 binding to gp120, but it does not interfere with the binding of other neutralizing or non-neutralizing monoclonal antibodies to gp120 (Esser et al., 1999). Conversely, 2G12 does not inhibit CV–N binding to gp120, probably because there are multiple binding sites for CV–N and only some of them are blocked by 2G12 (Esser et al., 1999). Both CV–N and 2G12 bind terminal α-Man-(1→2)-α-Man moieties on gp120. However, CV–N seems to bind a single N-linked glycan, whereas 2G12 interacts with two glycans at a time (Calarese et al., 2003; Chang and Bewley, 2002).

Molecular characterization of the 2G12 epitope contributes to the design of immunogens able to elicit 2G12-like antibodies. In one study, a novel template-assembled oligomannose cluster was synthesized as an epitope mimic for 2G12. Three $Man_9GlcNAc_2Asn$ moieties were selectively attached to a cholic acid scaffold. This synthetic cluster was a 46-fold more effective inhibitor than the micromolar 2G12 inhibitor, $Man_9GlcNAc_2Asn$. The binding affinity of $Man_9GlcNAc_2Asn$ is still an order of magnitude weaker than that of gp120. To increase specificity, new 2G12 mimics could be designed by optimizing the length of spacers and the rigidity of the scaffold molecule. This approach may lead to the development of carbohydrate-based vaccines against HIV-1 (Li and Wang, 2004).

4.2. Mannose-binding lectin (MBL)

Members of the Ca^{2+}-dependent (C-type) lectin family contain homologous carbohydrate-recognition domains (CRDs) that mediate a variety of interactions by binding selectively to different carbohydrates (Drickamer and Taylor, 1993). The C-type lectin family can be further subdivided into two subgroups, formed by the soluble collectins and by the cell-surface macrophage mannose receptor. Collectins have an oligomeric ''bouquet'' structure containing two to six building blocks. Each block consists of a polypeptide homotrimer. Serum mannose-binding proteins (rat MBP–A and human MBP) are hexameric with trimeric building blocks (see Fig. 5C). Liver MBP–C is a smaller oligomer and probably consists of two associated trimers (Mizuno et al., 1981). Serum Mannose-binding lectins (MBLs) act in the antibody-independent host defense by binding with high affinity to repetitive mannose and/or *N*-acetylglucosamine residues on the surface of microorganisms such as *Candida albicans*, *Salmonella typhimurium* and *Neisseria gonorrhoea* (van Emmerik et al., 1994). This binding triggers opsonization and complement-mediated cell lysis (Ikeda et al., 1987; Jack et al., 2001; Petersen et al., 2001; Weis et al., 1998). MBL-associated proteases MASP-1 and MASP-2 activate complement components C4 and C2 and catalyze the next step in the classical complement cascade (Thiel et al., 1997). Liver MBP–C may have a role in host defense, because it is structurally very similar to serum collectins.

MBLs also bind to the gp120 glycoprotein of HIV-1 (Haurum et al., 1993). HIV particles lacking gp120/gp41 do not bind MBLs, suggesting that carbohydrates present on gp120 mediate the interaction between whole virus and MBL (Saifuddin et al., 2000). More recently, it was shown that the MBL–gp120 interaction occurs via high-mannose carbohydrates found on the surface of gp120 (Hart et al., 2003).

There are four distinct regions in a typical MBL structure. The cysteine-rich amino terminal domain forms interchain disulfide bonds that stabilize the trimer. The amino terminal domain is followed by 18–20 Gly-X–Y repeats, a ''neck'' domain of ~30 amino acids, and a CRD domain of ~120 amino acids. The terminal β-strands of the domain are positioned close to each other and permit the CRDs to be located anywhere in a polypeptide chain. A structure-based sequence alignment of C-type lectin CRDs highlighted the essential elements of a CRD fold (Weis et al., 1991). Two pairs of conserved disulfide bonds stabilize the compactly folded structure and a set of conserved residues forms the calcium binding sites. The two calcium ions stabilize a series of loops and are involved in carbohydrate binding (Mullin et al., 1994; Weis et al., 1991).

The carbohydrate specificity of MBLs is fairly broad. MBLs recognize *D*-mannose, *N*-acetylglucosamine, and *L*-fucose. A common motif between these sugars is defined by the

presence of vicinal, equatorial, hydroxyl groups found in positions 3 and 4 of the sugar ring. MBLs do not bind with high affinity galactosides and sialic acids, which have an axial 4-hydroxyl group, and therefore an unfavorable stereochemistry (Weis et al., 1992). However, MBP–C can bind 1.3 M galactose in a way similar to that of other monosaccharides (Ng et al., 1996). While both MBP–A and MBP–C bind sugars that contain (+)-*syn*-clinal hydroxyl groups with similar affinity, MBP–A binds GlcNAc–BSA much more tightly than MBP–C does, and in general, the two MBP proteins have strikingly different binding specificities for complex carbohydrates and neoglycoproteins (Childs et al., 1990).

High-resolution crystal structures of rat serum mannose-binding protein A (MBP–A) and liver-associated MBP–C complexed with carbohydrates revealed the mode of carbohydrate binding (Ng et al., 1996, 2002; Weis et al., 1992). Most of the binding affinity is derived from direct coordination of the principal Ca^{2+} ion, with the 3- and 4-hydroxyl groups of the carbohydrate. Each hydroxyl group contributes a lone pair of electrons to enable the coordination of calcium. The other lone pair is a hydrogen bond acceptor from the NH_2 groups of two Asn residues. The Ca^{2+} atom is coordinated in a pentagonal bipyramidal geometry, in an intimately linked ternary complex of protein, Ca^{2+} and sugar (Ng et al., 1996).

Only the carbohydrate C4–Cβ His189 contact is energetically significant for sugar binding by MBP–A (Iobst et al., 1994), whereas an equivalent contact is provided by the Cβ of Val in a MBP–C complex. The small number of interactions between these proteins and carbohydrates reflects their high specificity for a broad range of carbohydrates. Since the mode of MBP–sugar interaction is symmetrical, binding is possible when the position of 3- and 4-hydroxyl groups is reversed (Ng et al., 1996; Weis et al., 1992). The different specificity may be attributed to the orientation of the bound sugar or to the binding of the oligosaccharide to another site, distinct from the Ca^{2+} site 2. A site similar to this one was identified in MBP–C in the presence of 1.3 M α-Me-Man (Ng et al., 1996). It is possible that this site forms part of an extended binding site that possesses high affinity only for ligands larger than a monosaccharide. A terminal Man, upon binding to MBP–A, buries 90 $Å^2$ of protein-accessible surface area (Weis et al., 1992), whereas α-Me-Man buries 80 $Å^2$ of MBP–C (Ng et al., 1996).

The orientation of monosaccharides bound to MBP–C is reversed relative to those bound to MBP–A. Despite the asymmetry of the monosaccharide, two orientations can be generated by rotation of the saccharide ring around the local two-fold symmetry axis. This rotation relates the equatorial 3- and 4-OH groups of mannose or their stereochemical equivalents. The binding orientation observed in MBP–A–$Man_6GlcNAc_2Asn$ was defined as orientation I, and the reverse as orientation II (see Fig. 5D). In orientation I, the 3-OH of mannose or its equivalent is bound to Glu185 or Asn187, whereas in orientation II, the 4-OH occupies this position (Ng et al., 2002). In the sugar-free MBP–C structure, two water molecules occupy the positions taken by the 3- and 4-OH groups of mannose. Interactions of the α-anomer with the His189 ring contribute to the binding energy and possibly lead to discrimination between α and β anomers (Ng et al., 2002).

MBP–C binds oligosaccharides in orientation II, whereas MBP–A binds these carbohydrates in orientations I and II (see Table 3). These differences in binding can be attributed to the less restrictive nature of the MBP–C site, which might more easily accommodate the non-terminal portions of these oligosaccharides.

Because MBPs do not show significant differences in their binding energies when they bind mono- or oligosaccharides, one might ask, how important is the ligand orientation by the binding

Table 3
Sugar-binding orientation at the principal Ca^{2+} site in mannose-binding proteins

Protein	Saccharide	Binding mode
MBP–A	Monosaccharides	II
	Disaccharide: α-Man-(1→3)-Man	I
	$Man_6GlcNAc_2Asn$: α-Man-(1→3)-Man	I
	$Man_6GlcNAc_2Asn$: α-Man-(1→2)-Man	I
	$Man_6GlcNAc_2Asn$: α-Man-(1→6)-Man	II
MBP–A (H189V)	Disaccharide: α-Man-(1→3)-Man	A: II B: II C: II
MBP–A (H189V/I207V)	Disaccharide: α-Man-(1→3)-Man	A: I (50%) II (50%) B: I C: I (50%) II (50%)
MBP–C	Monosaccharides	II
	Disaccharide: α-Man-(1→3)-Man	II
	Trimannosyl core	II
	High-affinity linear trimannose	II
	GlcNAc-terminated core	II
	Bivalent Man-terminated glycopeptide	II

Adapted from Ng et al. (1996, 2002). Estimated occupancies are for the protomers: A, B, C.

site? On one hand, the small binding energy differences observed in monovalent binding could compound to significant differences in multivalent binding. On the other hand, the specificity for a particular binding orientation might affect the ability of an oligomeric lectin to interact simultaneously with different branches of an oligosaccharide (Ng et al., 2002).

This high specificity to a broad range of terminal monosaccharides raises the question of why MBLs do not bind to the glycoproteins naturally present in host cells. First, vertebrate glycoproteins rarely contain terminal mannose, GlcNAc, or fucose residues. Second, the oligomeric structure of MBLs is likely to provide the specificity required to distinguish foreign from self. Similarly to other lectins, the dissociation constants of MBLs with carbohydrates are in the 2 mM range (Weis and Drickamer, 1996). Multivalent binding enhances the overall binding energy, which is the sum of the individual binding energies. Crystal structures of the MBLs also provide an explanation for their ability to distinguish foreign carbohydrates from self. A hydrophobic interface between the neck domains and the CRDs positions the CRD binding sites 53 and 45 Å apart, in the rat and human trimers, respectively. In contrast, terminal mannose residues in vertebrates are located only ~20–30 Å apart (Lee et al., 1991, 1992). The specificity of MBLs is largely due to their oligomeric structure and not their individual binding sites (Ng et al., 2002; Weis and Drickamer, 1994).

The CRDs of DC-SIGN and DC-SIGNR (see below) share 24% sequence identity with rat serum MBP–A, but they exhibit different oligosaccharide binding. In MBLs each CRD interacts with a terminal mannose or a GlcNAc residue in a ligand, and the rest of the oligosaccharide points away from the surface of the protein (Weis et al., 1992). By contrast, DC-SIGN and

DC-SIGNR present high affinity binding to internal features of high-mannose oligosaccharides (Feinberg et al., 2001).

4.3. DC-SIGN and DC-SIGNR

The primary immune response is typically initiated by the interaction of resting T cells with antigen-presenting dendritic cells (Banchereau and Steinman, 1998). T cell interaction with dendritic cells is mediated by the binding of the T cell surface receptor ICAM-3 to a dendritic cell surface receptor called DC-SIGN (Dendritic Cell-Specific Intercellular adhesion molecule (ICAM) Grabbing Nonintegrin) (Geijtenbeek et al., 2000c). DC-SIGN may participate in the trafficking of dendritic cells from the blood to the lymphatic system, via interactions with ICAM-2 (Geijtenbeek et al., 2000a).

DC-SIGN belongs to the C-type (Ca^{2+}-dependent) lectin superfamily and is a type II membrane protein with an extracellular domain that mediates tetramerization (Mitchell et al., 2001) and a C-terminal carbohydrate recognition domain (CRD) (Weis et al., 1998). Antibodies to DC-SIGN, Ca^{2+} chelators, mannan or mannose can inhibit T cell activation by dendritic cells (Geijtenbeek et al., 2000c). These results suggest that the interaction with ICAM-3 is via Ca^{2+}-dependent carbohydrate binding.

DC-SIGNR (DC-SIGN-Related) is a receptor with 77% sequence identity to DC-SIGN. DC-SIGNR is expressed on the endothelial cells of liver sinuses, lymph node sinuses and placental villi (Bashirova et al., 2001; Pohlmann et al., 2001c).

Further details regarding the specificity and mechanism of carbohydrate recognition by these lectins were provided by the crystal structures of CRDs from DC-SIGN and DC-SIGNR bound to pentasaccharides (Feinberg et al., 2001) (see Fig. 6A). The DC-SIGN and DC-SIGNR CRDs adopt a typical long-form C-type lectin fold (Drickamer, 1999). The unusually long second α-helix is more tightly packed against the domain CRD, whereas the C-terminal end of the CRD, together with a loop, forms a continuous surface that interacts with the carbohydrate. The bound pentasaccharide is very well defined and has virtually the same conformation in both structures. Both lectins interact with the same low-energy oligosaccharide conformation. The internal $\alpha 1 \rightarrow 3$-linked mannose (Man2) binds to the principal Ca^{2+} site, typical to CRDs. The equatorial 3- and 4-hydroxyls coordinate the Ca^{2+} ion and form hydrogen bonds with the amino acids that coordinate the Ca^{2+} ion. However, the interactions between the carbohydrate and the Ca^{2+} site are unusual because they involve a core carbohydrate rather than a terminal one. In DC-SIGNR-sugar binding, the exocyclic C6 atom interacts with Glu366, and the 6-OH interacts with Asn379 via a water molecule (Asn379 in DC-SIGNR is equivalent to Asp367 in DC-SIGN). The $\alpha 1 \rightarrow 6$-linked branch with two mannose residues contacts the solvent exposed Phe325 and Ser372. The phenylalanine side chain forms van der Waals interactions with the Man3 in the crevice that is established by Man3 and Man4. Next to Phe325, Ser372 hydrogen bonds to the 3-OH of Man4 and participates in a water-mediated hydrogen bond with the 2-OH of Man3. The backbone amide of Ser374 also forms hydrogen bonds to the 3-OH of Man4 via a water molecule (Feinberg et al., 2001).

The terminal GlcNAc residues form several contacts with DC-SIGN and DC-SIGNR. For DC-SIGNR, the GlcNAc1 on the $\alpha 1 \rightarrow 3$ branch is involved in hydrogen bonding with Ser363, whereas for DC-SIGN, the same GlcNAc1 participates in van der Waals interactions with a corresponding

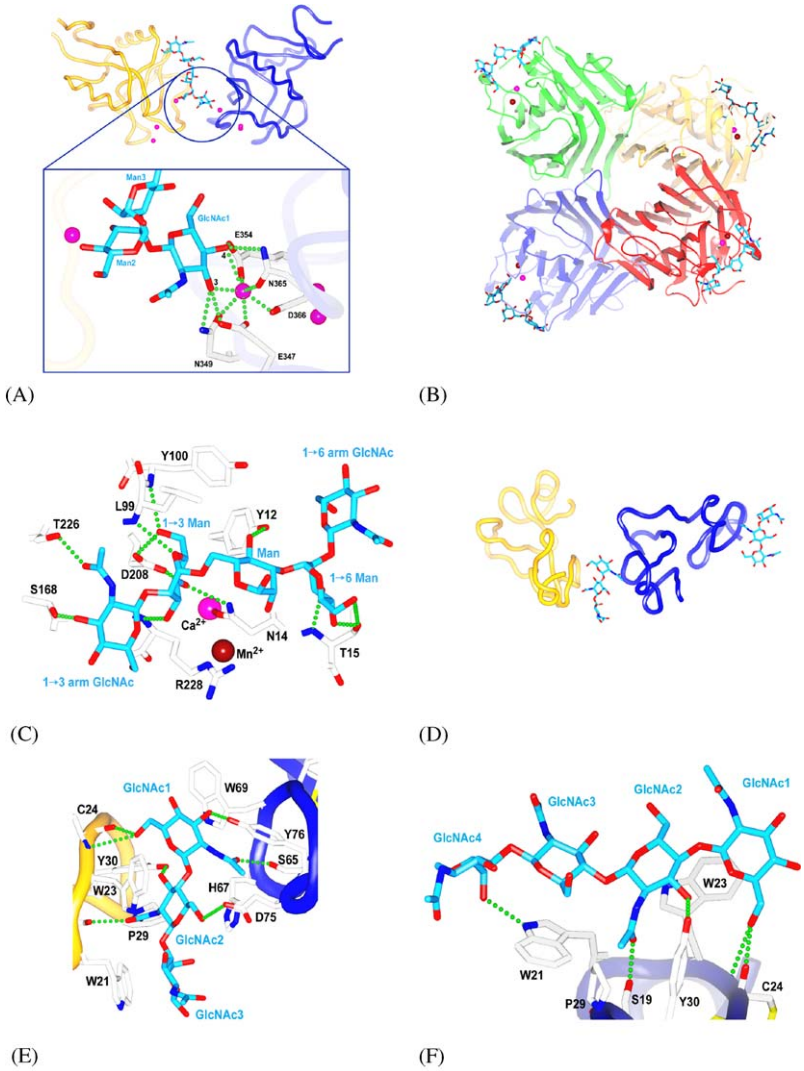

Fig. 6. (A) Ribbon diagram of DC-SIGN complexed with $Man_3GlcNAc_2$ (pdb code 1k9i, Feinberg et al., 2001). Ca^{2+} is shown as a purple sphere while the carbohydrate is shown in a blue stick representation. The Ca^{2+} coordination is shown in detail, with the hydrogen bonds to important coordinating residues shown as green spheres. (B) Ribbon diagram of the concanavalin A structure complexed to a pentasaccharide (pdb code 1tei, Moothoo and Naismith, 1998). The monomers are shown in different colors and the carbohydrates are in light-blue stick representation. (C) Hydrogen bonding of carbohydrate-binding residues in the structure of ConA (1tei, Moothoo and Naismith, 1998). The pentasaccharide βGlcNAc-(1→2)-α-Man-(1→3)-[βGlcNAc-(1→2)-α-Man-(1→6)]-Man is shown in light-blue stick representation and the hydrogen bonds as green spheres. The reducing mannose (see Table 4) is labeled Man. (D) Ribbon diagram of *Urtica dioica* agglutinin isolectin VI (pdb code 1ehh, Harata and Muraki (2000)). One carbohydrate molecule is bound to binding site I, while the second one binds to both sites I and II from two different monomers. (E) Hydrogen bonding and van der Waals interactions of $GlcNAc_3$ with UDA (1ehh, Harata and Muraki (2000)). The hydrogen bonds are shown as green spheres. (F) Hydrogen bonding and van der Waals interactions of $GlcNAc_4$ with UDA (1en2, Saul et al., 2000).

valine. This particular contact represents the only major difference in the interaction of the two proteins with the pentasaccharide. In the DC-SIGN crystal structure, GlcNAc1 also forms a typical CRD-type Ca^{2+} coordination with the principal Ca^{2+} site on the second molecule of the DC-SIGN dimer that crosslinks the two monomers. GlcNAc5 from the other carbohydrate branch forms a series of van der Waals interactions and hydrogen bonds with a pocket formed by Asn323 and Phe325 (Feinberg et al., 2001).

DC-SIGN and DC-SIGNR can bind to multiple strains of HIV-1, HIV-2, SIV and pass these viruses to both T cell lines and human peripheral blood mononuclear cells (Pohlmann et al., 2001c), playing a role in both horizontal and vertical transmission, as well as dissemination of virus within the host (Pohlmann et al., 2001a). DC-SIGN can also bind to laboratory-adapted X4 isolates of HIV-1 (Pohlmann et al., 2001b). Similar to the dendritic cell–T cell interaction, Ca^{2+} chelators and mannan can inhibit DC-SIGN and DC-SIGNR interactions with HIV (Bashirova et al., 2001; Geijtenbeek et al., 2000b; Pohlmann et al., 2001c). Gp120 harbors large clusters of high mannose oligosaccharides (Mizuochi et al., 1988a), a type of structure with which both DC-SIGN and DC-SIGNR interact efficiently (Mitchell et al., 2001). Quantitative inhibition assays of DC-SIGN and DC-SIGNR showed that the pentasaccharide used for the crystal structures had a comparable affinity to Man-5, while neither of the proteins bound complex oligosaccharides attached to glycoproteins (Feinberg et al., 2001). These experiments showed that DC-SIGN and DC-SIGNR are specific for high mannose N-linked oligosaccharides. Modeling of Man-9 structures from NMR experiments (Woods et al., 1998) into the DC-SIGN and DC-SIGNR structures revealed that, in the context of N-linked glycans, CRDs recognize the trisaccharide α-Man-(1→3)-[α-Man-(1→6)]-Man only with the central mannose linked in the α anomeric configuration. Phe325 has a dual role of providing surface shape complementarity to the α-Man-(1→6)-Man moiety and selecting between the inner and outer trimannose structures by preventing binding to the inner branch point mannose (Feinberg et al., 2001).

DC-SIGN- and DC-SIGNR-gp120 interactions represent a potential target for anti-HIV therapy aimed at preventing the lectin–virus interaction, aimed at lowering of the efficiency of T cell infection. The oligosaccharide binding mechanism of these lectins provides a starting point for the design of new therapeutics, which would interfere with the viral infection at a novel stage and could be potentially prophylactic (Feinberg et al., 2001).

4.4. Concanavalin A (ConA)

The carbohydrate-binding protein concanavalin A (ConA), isolated over 80 years ago by Sumner (1919) from the jack bean (*Canavalia ensiformis*), is one of the most studied lectins. ConA is a member of a larger family of plant lectins (Sharon, 1993). In the tetrameric protein, each 25 kDa monomer has a saccharide-binding site and two distinct metal-binding sites: one for a transition-metal ion and another one for a calcium ion (Kalb and Levitzki, 1968). ConA dissociates into a dimer below pH 6.5, losing the naturally occurring metals (at pH 1.2) and no longer binding saccharides. Addition of the missing metals at pH 5.2 restores saccharide binding (Yariv et al., 1968). ConA binds Man and Glc and is most specific for α-Man-(1→3)-[α-Man-(1→6)]-Man cores of N-linked oligosaccharides (Mandal et al., 1994).

As of the time of this writing, 47 crystal structures of ConA have been deposited into the Protein Data Bank (PDB). The structures range from ultra-high resolution (0.94 Å) (Deacon

et al., 1997) to various crystal forms (Goel et al., 2001; Hardman and Ainsworth, 1972; Harrop et al., 1996; Kantardjieff et al., 2002; Naismith et al., 1994; Shoham et al., 1979), unliganded or liganded (Bouckaert et al., 1999; Bradbrook et al., 1998; Moothoo and Naismith, 1998), and with or without bound metal ions (Bouckaert et al., 1995, 1996). From a series of structures determined for conA–saccharide complexes, several show limitations in revealing the exact binding mode, due to insufficient resolution or local disorder induced by sugar binding.

The topology of the 237-residue conA is mainly unchanged in all native and liganded crystal structures; saccharide binding does not involve significant conformational changes. The main structural feature of conA is a jellyroll motif formed by two large β-sheets that contain almost half of the protein residues (see Fig. 6B). Two dimers (monomers AB and CD) form a tetramer, with a 12 stranded antiparallel β-sheet on each face of the dimer (Reeke et al., 1975). The formation of the A–B and C–D dimers involves extensive interactions between the monomers, with the burial of 2300–2500 $Å^2$ of molecular surface. In the tetramer, monomer-to-monomer interactions AB, AC, CD and DB contribute equally to a total of 9600 $Å^2$ of buried surface area (Naismith et al., 1994).

Each monomer has a carbohydrate-binding site located in a shallow pocket near the surface of the protein, formed by residues Tyr12, Asn14, Leu99, Tyr100, Asp208 and Arg228. The saccharide–protein interactions are very similar for the four monomers, α-Man–OMe burying ~270 $Å^2$ of protein surface upon binding (Naismith et al., 1994). α-Man–OMe 3-OH, 5-OH, and 6-OH make hydrogen bonds to the backbone amide N atoms of Arg228, Leu99 and Tyr100, respectively. 6-OH also hydrogen bonds to the side chain of Asp208, while 4-OH hydrogen bonds the Asp208 and Asn14 side chains. Extensive van der Waals contacts are formed between 3-OH, 4-OH, C6 and 6-OH with the surface of the binding pocket (see Table 4).

An extensive solvent network is present in the sugar-binding sites. High-resolution structures reveal a water molecule above the ring of Tyr12, in a site resembling the C6 position of a bound carbohydrate (Deacon et al., 1997). Another water molecule implicated in ligand binding is hydrogen bonding to the amino groups of Asn14 and Arg228.

Table 4
Contacts in concanavalin A—carbohydrate complexes

	H-bonds	Polar contacts	Van der Waals contacts
Pentasaccharide complex			
1→6 arm GlcNAc	2	2	25
1→6 Man	6	2	44
Reducing Man	1	1	6
1→3 Man	3	2	23
1→3 arm GlcNAc	0	0	7
Trisaccharide complex			
1→6 Man	6	2	46
Reducing Man	1	1	8
1→3 Man	3	2	23
α-Man–OMe complex			
α-Man–OMe	6	2	48

Adapted from Moothoo and Naismith (1998). The sugar at the monosaccharide binding site is shown in bold.

Crystal structures of conA with larger saccharides show a binding mode similar to that of a monosaccharide (see Table 4). In complexes with the linear trimannoside (Naismith and Field, 1996) and pentasaccharide (Moothoo and Naismith, 1998), all carbohydrate rings are involved in direct contacts with the protein (see Fig. 6C). The 1→6 Man bound in the monosaccharide site participates in identical hydrogen-bonding and van der Waals interactions as α-Man–OMe, probably anchoring the other sugar residues (Moothoo and Naismith, 1998).

The transition-metal ion is located at an average distance of 12.5 Å from the center of the binding pocket, while the distance of the Ca^{2+} ion is 8.2 Å. The two metal-ion sites S1 and S2 are located 4.2–4.6 Å apart. The S1 site is generally occupied by Mn^{2+}, with an octahedral coordination, but Ni^{2+}, Zn^{2+}, Co^{2+} and Cd^{2+} can also be bound to this site. Site S2 harbors a Ca^{2+} ion in a pseudo-octahedral coordination. The metal ions play a structural role by maintaining saccharide-binding residues in their proper conformation (Bouckaert et al., 1995). The Ca^{2+} ion interacts with a loop of the binding pocket (Asp10, Tyr12, Asn14, Asp19), possibly stabilizing its geometry. In the absence of metals, sugar binding is abolished. Structures of metal-free conA reveal large conformational changes in the metal-binding regions and, implicitly, for the residues involved in saccharide binding (Reeke et al., 1978). The presence of the Ca^{2+} ion is critical, since a water molecule interacting with the Ca^{2+} ion is thought to stabilize the Ala207–Asp208 *cis* peptide bond (Naismith et al., 1994). In the absence of the Ca^{2+} ion the peptide bond undergoes *cis* to *trans* isomerization, followed by an expansion of the sugar-binding loop Leu99–Tyr100 (Bouckaert et al., 1995; Reeke et al., 1978). There seems to be an order in which metal ions bind to demetallized conA. The S1 site has to be occupied first, before binding of Ca^{2+} can occur in site S2. It is possible to saturate both sites with Ca^{2+} ions, although the binding of Ca^{2+} in S1 is weak and the ion can be easily replaced by Mn^{2+} (Brewer et al., 1983).

4.5. *Urtica dioica agglutinin (UDA)*

The stinging nettle (*Urtica dioica*) agglutinin isolated from rhizomes is an 8.5 kDa monomeric protein that specifically binds oligomers of *N*-acetylglucosamine (GlcNAc) (Peumans et al., 1984; Shibuya et al., 1986). UDA inhibits HIV-1, HIV-2, CMV, RSV, and influenza A virus-induced cytopathicity at EC_{50} of 0.3–9 μg/ml. UDA is also a potent inhibitor of syncytium formation between HIV-infected HUT-78 cells and $CD4^+$ MOLT-4 cells (Balzarini et al., 1992). The lectin is a superantigen for murine T cells, inducing the exclusive proliferation of $V\beta8.3^+$ lymphocytes (Galelli and Truffa-Bachi, 1993). It is unique among T cell superantigens in that it can be presented by both class I and II major histocompatibility complex (MHC)-bearing antigen presenting cells (Rovira et al., 1999). UDA is a very stable protein, capable of withstanding 0.1 N HCl, 1 N acetic acid or 5% TCA, as well as heating at 80 °C for 15 min. It gradually loses activity above pH 12. UDA has an extremely low agglutination activity (Peumans et al., 1984). The primary structure of UDA revealed two hevein-like domains with cysteine residues spaced identically as in hevein (Beintema and Peumans, 1992). Carbohydrate-binding experiments suggested that UDA has two binding sites with different affinities. UDA binds tetra-*N*-acetylchitotetraose ($GlcNAc_4$) three times stronger than $GlcNAc_3$ and binds $GlcNAc_3$ 30-fold stronger than $GlcNAc_2$ (Shibuya et al., 1986). UDA is a mixture of seven individual isolectins with identical agglutination and sugar-binding specificity (Van Damme et al., 1988a, b). Isolectins II and III have 90 and 80 residues, respectively, while the other isoforms contain 89 residues

(Harata and Muraki, 2000). None of the residues that differ between the isoforms participate directly in carbohydrate binding (Saul et al., 2000).

The crystal structure of UDA isolectin VI (UDA-VI) shows two domains: residues 1–42 and residues 47–89, analogous to hevein (see Fig. 6D). The molecule possesses little secondary structure with two short β-strands forming an antiparallel β-sheet in each domain, connected by several β-turns. Four disulfide bonds maintain the structure of each domain. The two domains are only 42% homologous, but their backbones superimpose well, except of some loop regions. Domain I is more similar in structure to the domain of wheat germ agglutinin (WGA) than to the second domain of UDA (Wright, 1977). The segment connecting the two UDA domains is four residues longer than the segments connecting the hevein domains in WGA (Saul et al., 2000). The two carbohydrate-binding sites are located at the ends of the dumbbell-shaped molecule (Harata and Muraki, 2000; Saul et al., 2000). A bound $GlcNAc_3$ molecule is sandwiched between the binding sites of domain I, and that of domain II from a symmetry-related molecule. Domain I has a carbohydrate-binding site formed by residues Ser19, Trp21, Trp23, and Tyr30. Trp21 and Tyr23 stack onto carbohydrate rings C and B, respectively, while Ser19 and Tyr30 form hydrogen bonds with the 3-OH group of ring B (see Fig. 6E). The binding site of domain II is formed by the homologous residues, Ser65, His67, Trp69 and Tyr76. Trp69 stacks with sugar ring A of $GlcNAc_3$, while His67 sandwiches sugar ring B, together with Trp23 from domain I of a different UDA molecule. Tyr76 hydrogen bonds with the 3-OH group of ring A (Harata et al., 2001). Of the $GlcNAc_3$ sugar residues, ring C exhibits the largest conformational flexibility. The binding site of UDA domain I interacts with all the sugar rings of both the trisaccharide and tetrasaccharide, while site II interacts only with sugar rings A and B of the former oligosaccharides (Harata and Muraki, 2000; Saul et al., 2000).

The stoichiometry of binding of UDA is one molecule of lectin to one molecule of carbohydrate. The affinity difference between the two sugar-binding sites can be attributed to the difference in amino acid residues involved in sugar binding. Trp21 from the carbohydrate-binding site of domain I is replaced by His67 in domain II. This confers a different hydrophobic character to the site in domain II, where the presence of two His residues also triggers greater pH dependency (Harata and Muraki, 2000). Domain I binding site can interact with any ring of the oligomeric GlcNAc, whereas the site in domain II is specific for the non-reducing terminus of the ligand. This implies that site I can attach anywhere on the chito-oligosaccharide chain. The sharing of a sugar molecule by neighboring UDA molecules is made possible by the extended conformation adopted by the GlcNAc oligomer (Saul et al., 2000).

UDA isolectin I differs by five amino acids from isolectin VI, but exhibits identical carbohydrate-binding activity (Van Damme et al., 1988b) (see Fig. 6F). Its crystal structure revealed a virtually identical backbone conformation and sugar-binding sites compared to those of UDA-VI. However, Zn^{2+} from the crystallization buffer was found to block the sugar-binding site of domain II, mediating a dimeric contact between domain II and I of a second UDA molecule. Although metal ion-mediated intermolecular interactions were observed for other dimers, in this case it is unlikely to have any physiological significance (Harata et al., 2001). Isolectin V, which differs from isoform VI by Gly9 instead of Ser, also has an identical backbone structure (Saul et al., 2000).

4.6. *Myrianthus holstii lectin (myrianthin, MHL)*

The roots of the African plant *Myrianthus holstii* contain a 9284 Da lectin myrianthin (MHL) that exhibits potent anti-HIV activity. MHL contains 16 disulfide-linked cysteine residues. Sequence analysis successfully assigned 79 amino acid residues out of 88, suggesting the presence of multiple isoforms differing in their primary structures at positions 52, 66 and 69 (see Fig-MHL). Sequence alignment of MHL with UDA (Beintema and Peumans, 1992; Chapot et al., 1986; Peumans et al., 1984) showed 61% homology between the two proteins. MHL also showed significant sequence homology with other plant lectins, such as wheat germ agglutinin, rice lectin, and chitinases from maize and tobacco (Charan et al., 2000).

CEM-SS cells were protected by MHL from the cytopathic effects of HIV-1_{RF}, with an EC_{50} value of 1.4 μg/ml (150 nM), an affinity similar to that of UDA (Balzarini et al., 1992). MHL did not prove to be toxic to the target cells even at the highest tested concentration (250 μg/ml). However, the anti-HIV activity was lost upon reductive cleavage of the disulfide bonds, emphasizing their structural importance. Virus particles treated with 10 μg/ml MHL, followed by dilution beyond effective concentrations of the protein, were fully infective, indicating that the protein was not virucidal. Further studies revealed that MHL acts reversibly to inhibit HIV infection of the host cells and its presence is continuously required for full activity (Charan et al., 2000).

Screening for potential molecular targets of MHL revealed that it binds to gp120. Unlike other plant proteins such as jacalin (Beintema and Peumans, 1992; Corbeau et al., 1994), MHL can bind gp120 simultaneously with soluble CD4, suggesting that MHL and CD4 bind gp120 at different sites. From a series of sugars tested, only GlcNAc interferes with MHL–gp120 binding. Once bound to GlcNAc, MHL is subsequently unable to bind gp120, implying that MHL interacts with carbohydrate moieties on the surface of gp120 (Charan et al., 2000).

MHL has lectin-like properties similar to those reported for UDA (Balzarini et al., 1992; Chapot et al., 1986; Peumans et al., 1984), but unlike other plant lectins MHL does not agglutinate human erythrocytes and does not show chitinase activity. For therapeutic use, a shortcoming of MHL, as that of other lectins, is their lack of specificity. Its affinity for a terminal GlcNAc monosaccharide is not as selective as binding to a higher-order oligosaccharide structure (Charan et al., 2000).

4.7. *Jacalin*

Jacalin is a tetrameric lectin from jackfruit (*Artocarpus integrifolia*) seeds. Jacalin shows 57% sequence identity with artocarpin from jackfruit seeds; however, it is a glycoprotein and is specific for galactose (Bourne et al., 2002). Each subunit is composed of a 133 residue α-chain and a 20 residue β-chain, produced by post-translational proteolysis (Aucouturier et al., 1987; Mahanta et al., 1992; Yang and Czapla, 1993). Jacalin exhibits antiviral properties against a few viral isolates (Corbeau et al., 1994), interacting with the oligosaccharides from gp120 (Corbeau et al., 1995). At the disaccharide level, jacalin is specific to the tumor-associated sugar Galβ1–3GalNAc, known as T-antigen (Sastry et al., 1986).

Several jacalin crystal structures are available, in complex with α-Gal–O–Me (Sankaranarayanan et al., 1996), Galβ1–3GalNAc (Jeyaprakash et al., 2002), Gal, α-GalNac–O–Me,

Galβ1–3GalNAc–α-O–Me (Me-α-T-antigen), GalNAcβ1–3Gal–α-O–Me and Galα1–6Glc (melli-biose) (Jeyaprakash et al., 2003). Each 16.5 kDa jacalin subunit has a three-fold symmetric β-prism fold, composed of three four-stranded β-sheets (see Fig. 7A). Two sheets have Greek-key topology, while the third one is Greek key-like with a break in the outer loop, caused by the post-translational cleavage. The carbohydrate binding site is formed by loops 46–52, 76–82, and 122–125 that connect the N-terminus of the α chain to the inner strands of the Greek keys. The primary binding site is formed by the side chains of Phe47, Tyr78 and Asp125, backbone nitrogen and oxygen atoms of Tyr122, Trp123, and the amino terminus of the α-chain. Secondary site A involves the side chains of Tyr78, Tyr122, and Trp123, while secondary site B contains the backbone nitrogen and oxygen atoms of Val79, OG of Ser119, and the carboxy terminal region of the β-chain. The primary binding site is occupied by Gal/GalNAc/Man with the anomeric carbon atom pointing towards the secondary site A. This geometry is favorable for α substituents; however, with β-substituted sugars, the reducing sugar binds at the primary site with the non-reducing end in the secondary site B (see Table 5) (Jeyaprakash et al., 2003).

In the case of a Gal ring bound in the primary pocket, the terminal amino group of the α chain hydrogen bonds to 3-OH and 4-OH of the sugar ring, Tyr122 N to 5-OH and 6-OH, the NH and O of Trp123 to 6-OH and the side chain oxygen atoms of Asp125 to 4-OH and 6-OH (see Fig. 7B). The Tyr78 ring stacks on the carbohydrate ring. An additional O-methyl substituent on the carbohydrate ring will interact with the Tyr122 ring, whereas an *N*-acetylamino substituent hydrogen bonds with the amino terminus of the α chain. Secondary site A binds methyl groups or sugar residues α-linked to the residue in the primary site. This α-linked group nestles among the side chains of Tyr78, Tyr122 and Trp123, participating mainly in hydrophobic interactions. Secondary site B shows considerable variability in its interactions with the carbohydrate (Jeyaprakash et al., 2003).

Modeling of a mannotriose into the binding site suggests that jacalin has restricted ability to bind N-linked glycans. This is also supported by binding data, which show a higher affinity of jacalin for O-glycans. Non-specific protein–protein interactions might provide a substantial contribution when jacalin binds glycoproteins containing high-mannose carbohydrates (Jeyaprakash et al., 2003).

4.8. *Galanthus nivalis agglutinin (GNA)*

Galanthus nivalis agglutinin (GNA), isolated from snowdrop bulbs, belongs to the monocotyledonous non-seed lectin superfamily (Van Damme et al., 1987). These lectins are highly homologous and are completely unrelated to the Glc/Man/Gal specific family of dicotyledonous legume lectins (Sharon and Lis, 1991) and the C-type mannose-binding animal lectins (Drickamer and Taylor, 1993). GNA agglutinates rabbit erythrocytes, but not human ones. Solution binding studies showed that GNA has exclusive specificity for oligosaccharides containing $\alpha 1 \rightarrow 3$ and $\alpha 1 \rightarrow 6$-linked mannose moieties (Shibuya et al., 1988) and some affinity for amylose, dextran, and glycogen (Van Damme et al., 1987). Their inhibitory effect on HIV infectivity during the fusion step is attributed to specific interactions with branched high-mannose carbohydrates on the viral gp120 (Balzarini et al., 1991).

GNA is tetrameric, with 12.5 kDa subunits that are not held together by disulfide bridges. The pure protein withstands a wide range of pH (3–12) and high temperatures (70 °C) for 10 min

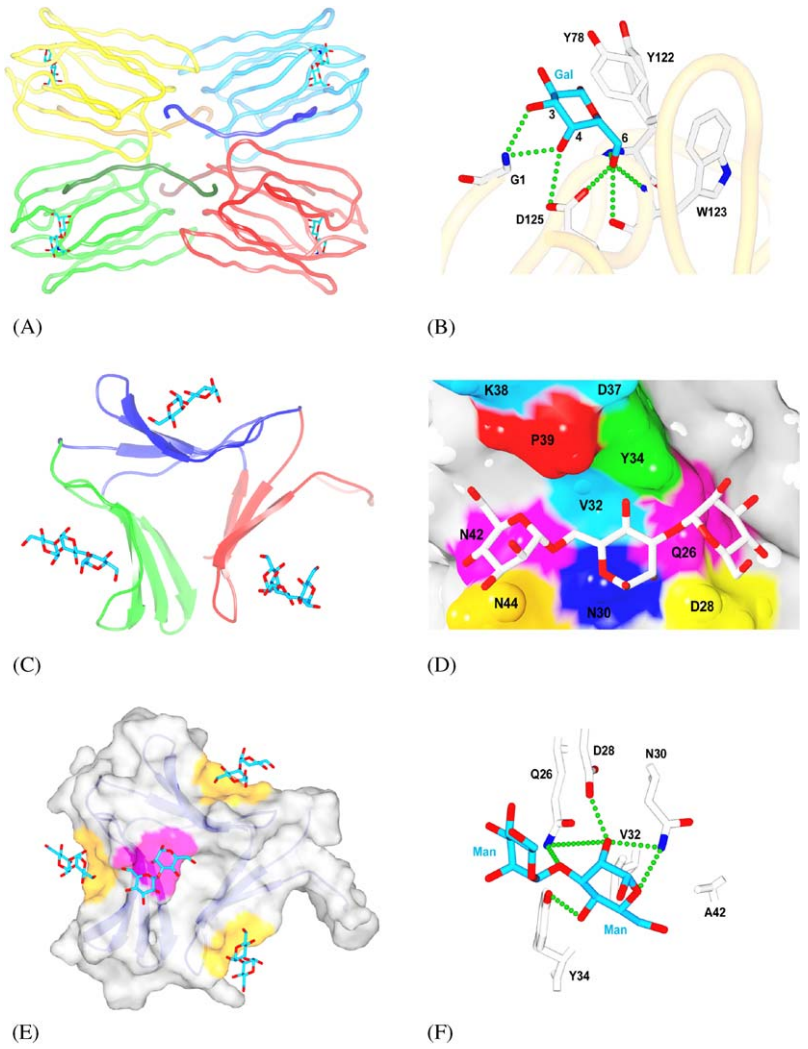

Fig. 7. (A) The β-prism structure of jackfruit seed jacalin (pdb code 1m26, Jeyaprakash et al., 2002) complexed with T-antigen (light-blue sticks). Each monomer (color coded) is composed of a large α-chain (light color) and a short β-chain (dark color). The sugar-binding sites are well-spaced from each other. (B) Galactose binding into the primary sugar-binding site of jacalin (1ugw, Jeyaprakash et al. (2003)). Only residues that interact with the Gal molecule (light-blue sticks) are shown. Hydrogen bonds are depicted as green spheres. (C) Ribbon diagram of GNA bound to a pentasaccharide (1jpc, Wright and Hester, 1996). The subunits shown in different colors form a typical ''monocot lectin consensus fold''. Carbohydrate residues are shown in light-blue stick representation. (D) Molecular surface of the carbohydrate-binding site of GNA (1jpc, Wright and Hester, 1996). The residues involved in binding are colored. (E) Molecular surface of NPL complexed to α-Man-(1$\rightarrow$3)-Man (ligh blue) (1npl, Sauerborn et al., 1999). The three typical carbohydrate binding sites are highlighted in yellow and the fourth one in purple. (F) Hydrogen bonding pattern of the NPL CRD complexed to α-Man-(1$\rightarrow$3)-Man (1npl, Sauerborn et al., 1999). The carbohydrate is shown in light-blue stick representation, with the hydrogen bonds depicted as green spheres.

Table 5
Mode of carbohydrate binding in different jacalin–sugar complexes

Bound saccharide	Primary site	Secondary site A	Secondary site B	Buried surface area (Å)	Binding ct. ($10^{-3} \times K_a$)
Gal	Gal	—	—	350	1.22
α-Gal–OMe	Gal	Me	—	390	40.0
α-GalNAc–OMe	GalNAc	Me	—	430	73.28
Galβ1–3GalNAc	GalNAc	—	Gal	477	122.0
Galβ1–3αGalNAc–OMe	GalNAC	Me	Gal	540	806.0
GalNAcβ1–3αGal–OMe	Gal	Me	GalNAc	523	302.0
Galα1–6Glc	Gal	Glc	—	468	5.31

Adapted from Jeyaprakash et al. (2003).

(Van Damme et al., 1987). GNA is synthesized as a 159-residue precursor with an N-terminal 23-residue signal peptide and a 29-residue C-terminal extension, both of which are removed post-translationally to yield the mature 105-residue protein (Van Damme et al., 1991). Six different isolectins have been identified, with sequence variation in the C-terminal region.

The crystal structure revealed a novel, β-barrel subunit fold, phylogenetically conserved among all members of the bulb lectin superfamily (see Fig. 7C). The β-barrel forms a non-polar core and is composed of three antiparallel β-sheets, which can serve as carbohydrate recognition domains, each with four β-strands (Hester et al., 1995). The lack of a β-sheet twist and the flat shape and sequential folding pattern of the β-sheets distinguishes the GNA motif from the β-trefoil fold. Each subunit (or CRD) contains a mannose-binding site. In the tetrameric GNA structure complexed with α-Man–OMe, the carbohydrate is bound with high occupancy in only one mannose-binding site in each subunit. This binding site is located in the C-terminal region of each subunit, and plays an important role in stabilizing the structure by inter-subunit strand exchange (Hester et al., 1995). The C-terminal strand exchange between subunits A–D and B–C generates 33 hydrogen bonds and buries 1738 $Å^2$ surface area, which is more than three times larger than between the non-interacting termini of subunits A–B and C–D (Hester and Wright, 1996). Unlike in concanavalin A (Naismith et al., 1994), the four sugar-binding sites of the tetramer are not spaced at maximum distance from each other (Hester et al., 1995). A side of the sugar ring (C2–C4) is nestled against a polar surface on one subunit, while the C-terminus of an opposing subunit partially encircles the saccharide. There are four strong hydrogen bonds between 2-OH and side chains of Asp91 and Asn93; 3-OH and Gln89; and 4-OH and Tyr97. Non-polar and polar interactions between ring atoms C3 and C4 with Val95, 1-O–Me and 6-OH with Asn83, Asp100 and His107 further immobilize the carbohydrate. In contrast with legume lectins and mammalian C-type lectins, GNA does not require the presence of metal ions or stacking interactions with aromatic side chains in the binding site. The five residues lining the polar surface of the binding pocket are well conserved in all three subdomains and throughout the lectin superfamily. Superposition of the 12 binding-site residues from the tetramer showed an identical topology, with r.m.s.d. of 0.23–0.7 Å. The specificity for Man can be attributed to the favorable conformation of the axial 2-OH group of the carbohydrate, participating in two hydrogen bonds, whereas an equatorial OH group, as in Glc, would be much less favorable (Hester et al., 1995).

GNA complexed with α-Man-(1→3)-α-Man–OMe has a different crystal form and shows different carbohydrate binding. The binding site in CRD1 is most specific for terminal non-reducing or reducing mannose, while CRD3 has an extended site, which is complementary to oligosaccharides with α1→3 linked mannose. GNA is the first plant lectin for which two carbohydrate binding modes are observed within the same subunit (see Table 6), suggesting that the protein takes advantage of multiple binding modes when it binds to complex oligomannan receptors (Hester and Wright, 1996).

The crystal structure of a complex with a branched mannopentaose (Man-5) revealed distinct binding only to CRD1 and CRD3 (see Fig. 7D). In one mode of binding the outer 3,6 tri-Man arm (M1–M2–M3) crosslinks two CRD1 pockets from twofold related GNA dimers. In this mode, only the terminal sugars are in contact with the protein surface. In the second binding mode, the same trimannoside is bound to an extended binding region of the CRD3 domain, which includes the conserved monosaccharide-binding pocket (Wright and Hester, 1996). M2 is positioned into the conserved monosaccharide pocket, the terminal M3 interacts with the same subsite as observed for the α-Man-(1→3)-α-Man–OMe complex (Hester and Wright, 1996), with its axial 2-OH hydrogen bonding to side chains of Asp37 and Lys38 from the two-fold-related subunit. The interfaces between subunits A–B and C–D have lower stability than those of the A–D or B–C dimer interfaces that form part of the CRD1 sites. M1 interacts with Pro39 and Ala42 that are part of a non-polar subsite located on the same subunit. There is an additional long hydrogen bond between 4-OH and the amide group of Asn44. In both binding modes, only the outer trimannoside branch is involved in binding interactions. For the CRD2 site the binding was weaker, with M2 bound in the conserved monosaccharide pocket and M1 oriented toward a non-polar subsite analogous to that in the CRD3 site. The M3 sugar residue is not visible, since it does not have a special binding region as in CRD3. These results further support the view that carbohydrates sterically adapt to the geometry of the protein surface while maintaining their low energy conformation in solution (Wright and Hester, 1996).

The 80–90% sequence identity between GNA and members of the bulb lectin superfamily (*Amaryllidaceae, Orchidaceae,* and *Alliaceae)* suggests the conservation of this structural fold between these proteins. Specificity differences between the family members might be attributed to the oligomeric state of these lectins (dimer or tetramer) and variability of the amino acid sequence. Precipitation studies with linear α1→3- and α1→6-linked mannose oligomers, showed that tetrameric HHA is precipitated by tetrasaccharides, while tetrameric GNA and dimeric NPA require oligomers of at least five residues in length for precipitation (Kaku and Goldstein, 1992).

In comparison with *Leguminoseae* lectins, *Amaryllidaceae* lectins bind α1→3- and α1→6-linked oligomannosides 10–50 times weaker and have 160–500 times lower association constants for

Table 6
Sugar-binding modes in *Galanthus nivalis* agglutinin (GNA) (Hester et al., 1995; Hester and Wright, 1996; Wright and Hester, 1996)

Saccharide	Binding mode	PDB code
α-Man–OMe	I	1msa
α-Man-(1→3)-Man–OMe	I, II	1niv
Man-5	I, II	1jpc

α-Man–OMe than ConA (Kaku et al., 1991). On the other hand, they show much greater propensity to form precipitates with linear mannans. The modulation of their mannose-binding affinity through inter-subunit contacts is also a unique property of these lectins, their monosaccharide pocket accommodating non-reducing terminal Man residues, internal Man residues, and reducing Man residues (Wright and Hester, 1996).

Binding studies suggested that lectins might destabilize the conformation of the gp120–gp41 complex (Favero, 1994). A possible mechanism by which gp120 binding to specific receptors is blocked might involve crosslinking of lectins with branched oligosaccharides (Dragic et al., 1996).

4.9. *Narcissus pseudonarcissus lectin (NPL)*

Narcissus pseudonarcissus lectin (NPL), isolated from daffodil bulbs, is a member of the monocot mannose-binding lectin family. NPL has high affinity for α1→3 or α1→6-linked mannoses (Van Damme et al., 1988a) and in agglutination assays is best inhibited by α1→6-linked mannotriose (Kaku et al., 1990; Shibuya et al., 1988). In solution, the protein is dimeric or trimeric, with ~12 kDa monomers. NPL is encoded by up to 200 genes with more than 50 different isoforms identified through isoelectric focusing, with comparable biochemical and carbohydrate-binding properties. Sequences of at least seven NPL isoforms can be distinguished on the level of cDNA (Van Damme et al., 1992; Van Damme and Peumanns, 1990). Similarly to other lectins from this family, NPL is expressed as a precursor with a 20–25 residue N-terminal leader sequence and a C-terminal tail that are cleaved to yield the 109 residue mature protein (Sauerborn et al., 1999).

The crystal structure of isoform 7 (NPL7) reveals a tetramer (Sauerborn et al., 1999) with typical "monocot lectin consensus fold" (Wright et al., 1997) (see Fig. 7E). Three four-stranded antiparallel β-sheets form a barrel with a three-fold axis at the center in each monomer. Each face of the triangular prism constitutes a CRD, with the essential residues for binding conserved (Sauerborn et al., 1999). Dimers are formed by strand-exchange of residues 102–109 from two 2-fold symmetry related monomers. This strand exchange is facilitated by the *cis* conformation of the peptide bond between Gly98 and Thr99, and the Trp102′ side chain reaching into the cleft formed by Trp41 and Trp73. Strand exchange buries 1850 $Å^2$ surface area. By bringing an extensive number of side chains together and shielding their hydrophobic patches, two dimers form a tetramer (Sauerborn et al., 1999). Their interface is equivalent to the A/B interface of GNA (Hester et al., 1995).

All three CRDs in an NPL7 monomer are equally occupied by a ligand. The β-hairpins between strands 2 and 3 of each β-sheet define a binding pocket. Three residues (Asn, Asp, and Gln) create a polar patch in the pit of the CRD pocket that strongly restricts the carbohydrate ligand to an axial hydroxyl at position 2 (see Fig. 7F). The position of the ligand in the pocket blocks access to Man O2 and O4, while O1, O3 and O6 are accessible to substituents. The main binding pocket can bind either a reducing residue (M1) or a non-reducing residue (M2) of an α1→3-linked oligomannose. CRD I binds the oligomannose M2, with M1 bridging to another molecule, while CRD II and CRD III bind M1, with M2 pointing into a large open chamber, depending on available charges. The three binding sites are equally occupied by carbohydrates, unlike in GNA. Calorimetry studies also show equal affinity for all three binding sites of NPL7. In addition to the three conserved CRDs, a unique fourth site was observed near the Trp cluster. CRD IV binds carbohydrates with low affinity in alternate conformations, with 0.25 occupancy for each

conformation, and this may explain the high affinity of this lectin for complex hybrid oligomannoses (Sauerborn et al., 1999).

Daffodil lectin binds to high-mannose or hybrid oligosaccharides present on the HIV-1 glycoproteins (Balzarini et al., 1991; Weiler et al., 1990). Hybrid glycans are bound with low affinity, possibly due to the presence of a GlcNAc at Man O4 (Kaku et al., 1990). Results of calorimetric studies indicate that $\alpha 1 \rightarrow 3$ mannobiose binding to NPL is exothermic, with a K_a of $500\,M^{-1}$. This suggests that strong binding can be achieved only by the simultaneous use of multiple binding sites to immobilize large oligosaccharides (Sauerborn et al., 1999).

4.10. *Scilla campanulata lectin (SCL)*

SCL isolated from bluebell (*Liliaceae* family) bulbs is a 119 residue, 13.1 kDa, mannose-specific lectin. Its anti-retroviral activity is an order of magnitude lower than that of lectins from the *Amaryllaceae* family (Wood et al., 1999). The SCA sequence is 81% homologous to HHA.

The crystal structure of the unliganded SCA tetramer revealed a typical monocot mannose-binding lectin fold (see Fig. 8A) where three β-sheet subdomains form a β-barrel-like arrangement with a pseudo-three-fold axis on the barrel axis. The consensus carbohydrate-binding sequence QXDXNXVXY present in *Amaryllaceae* lectins is present in CRDs I and III. CRD II contains two mutations, Asn to Val and Tyr to Ile, which probably prevent α-*D*-Man binding. Comparison of SCA subunits to GNA and HHA yielded r.m.s. deviations in the range 0.74–0.8 Å for main chain atoms. There is close agreement between the saccharide-binding residues of native SCA and the corresponding residues in GNA, HHA, and SCA complexes. These results suggest that interaction of these lectins with the oligosaccharides involves relatively small conformational changes of the binding residues (Wood et al., 1999).

The functional SCA is tetrameric in solution, with two different types of interfaces between the monomers. For the A–B dimer, C-terminal strand exchange gives rise to extensive contacts and a buried surface area about five times larger than in the A′–B or A–B′ dimers (Wood et al., 1999).

4.11. *Other monocot family lectins*

The tetrameric HHA from amaryllis (*Hippeastrum hybrid*) bulbs is another lectin from the monocot mannose-binding lectin superfamily. The crystal structure of its unliganded form shows a typical fold characteristic for this family (Chantalat et al., 1996).

LOA isolated from leaves of twayblade (*Listera ovata, Orchideaceae*) is another mannose-specific lectin. LOA has high affinity for $\alpha 1 \rightarrow 3$-linked Man oligomers (Balzarini et al., 1991). The $\alpha 1 \rightarrow 3$- and $\alpha 1 \rightarrow 6$-Man-specific lectins are potent inhibitors of HIV replication in vitro and they suppress giant cell formation between HUT-78/HIV-1 cells and uninfected MOLT-4 cells. The inhibitory effect is due to interaction with an early stage of the virus replicative cycle. In this respect, these lectins have a similar action to sulfated polysaccharides, but they do not inhibit virus binding to $CD4^+$ target cells. Unlike sulfated polysaccharides, these lectins are only inhibitory to HIV and CMV, with strong inhibition of giant cell formation. These lectins, applied together with sulfated polysaccharides, have a strong synergistic activity against HIV. They can also block virus entry into the cell, even if they are added 4–8 h postinfection. These results suggest that plant lectins interfere with the fusion process (Balzarini et al., 1991).

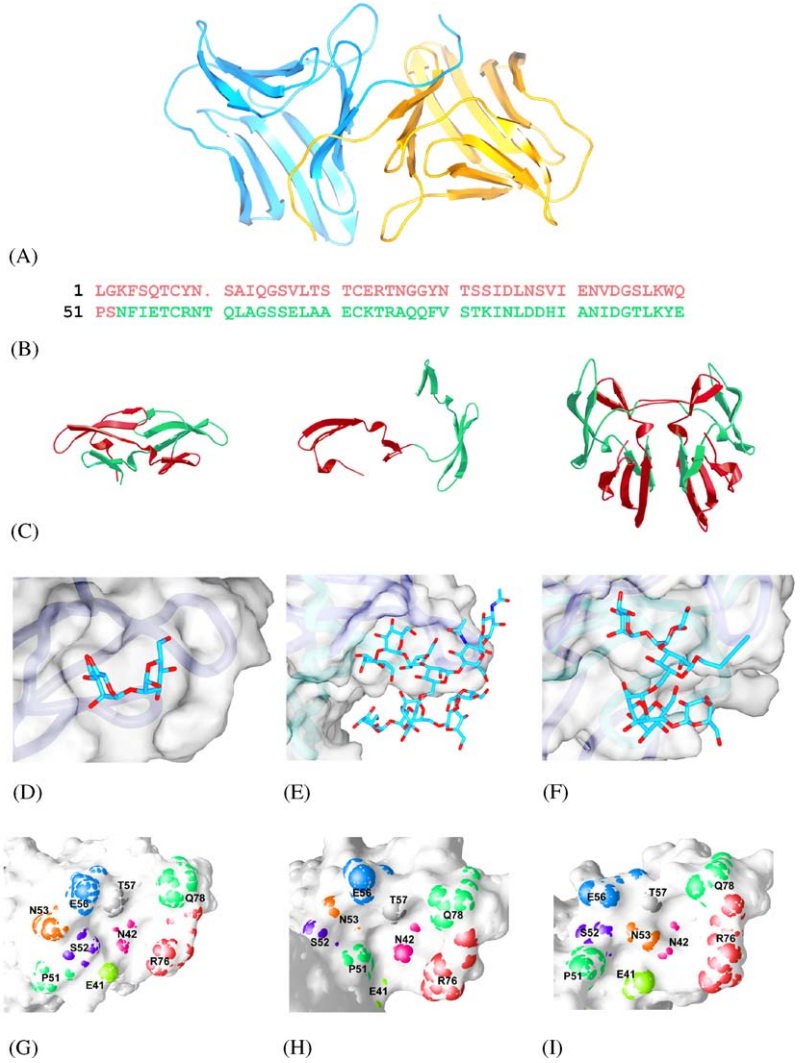

Fig. 8. (A) Ribbon diagram of *Scilla campanulata* lectin dimer (pdb code 1b2p, Wood et al., 1999). The strand exchange between the monomer termini is illustrated. (B) Sequence alignment of the two CV–N domains, with the domains color-coded as in panel C. (C) Structural forms of CV–N: compact monomer (2ezm, Bewley et al., 1998), domain-swapped extended monomer (3ezm, Yang et al., 1999) and domain-swapped dimer (1l5b, Barrientos et al., 2002a). Panel C generated with programs Bobscript (Esnouf, 1997) and Raster3D (Meritt and Murphy, 1994). (D) Structure of CV–N complexed to oligosaccharides: α-Man(1→2)-Man dimannose (1iiy, Bewley (2001)), (E) Man-9 (1m5m, Botos et al. (2002b)), and (F) synthetic hexamannose (1m5j, Botos et al. (2002b)). The backbone worms of the monomers (dark blue and cyan) are visible through the transparent molecular surface. (G) Oligosaccharide-binding pocket architecture, with residues involved in protein–ligand interaction numbered. Compact solution-structure monomer (1iiy, Bewley (2001)); (H) Trigonal crystal structure (3ezm, Yang et al., 1999); (I) Tetragonal crystal structure (1l5b, Barrientos et al., 2002a). Panels G-I generated with program GRASP (Nicholls, 1992).

4.12. Cyanovirin–N

The proteinaceous constituents of natural product extracts represent a largely untapped source of potentially novel, biologically active molecules (O'Keefe, 2001). Cyanovirin–N (CV–N), an 11 kDa protein from the cyanobacterium (blue-green algae) *Nostoc ellipsosporum* (Boyd et al., 1997) is a highly potent virucidal agent that has generated interest as a lead natural product for the prevention and chemotherapy of HIV infection (Botos and Wlodawer, 2003; De Clercq, 2000). CV–N has no significant homology (less than 20%) to any other known protein. The sequence shows a high degree of internal duplication between part A (residues 1–50), and part B (51–101) (see Fig. 8B). 16 residues (32%) are identical and 13 residues (26%) are homologous, indicating a probable result of gene duplication during evolution (Bergsma et al., 1975; Gustafson et al., 1997). The loops defined by the disulfides Cys8–Cys22 and Cys58–Cys73, and the region farthest from these loops (residues 40–49 in part A, and 91–100 in B) show the highest degree of homology. CV–N is stable and active after several freeze-thaw cycles, as well as after treatment with organic solvents (CH_3CN, CH_3OH, DMSO), denaturants (8 M GnHCl), detergents (0.5% SDS), 0.5% H_2O_2, or even 15 min boiling (Boyd et al., 1997).

Both natural and recombinant CV–N (Mori et al., 1998) are highly potent virucidal agents. They irreversibly inactivate diverse T-lymphocyte-tropic, macrophage-tropic and T-lymphocyte- and macrophage-tropic primary isolates of HIV-1, HIV-2, simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), and certain other enveloped viruses (Boyd et al., 1997; Dey et al., 2000). CV–N inhibits in vitro fusion of HIV-infected and noninfected cells and cell-to-cell transmission of HIV-1 infection (Boyd et al., 1997). CV–N association with gp120 is necessary but not sufficient for its antiviral activity (Mori et al., 1997a). CV–N is not lethal to the host cells at the concentration of 45–400 nM, with slight effect on viable cell numbers at 9000 nM (Boyd et al., 1997). Pretreatment with soluble gp41 ($IC_{50} = 24\,\mu M$) and soluble gp120 ($IC_{50} = 0.9\,\mu M$) completely abrogates the anti-HIV activity of CV–N (O'Keefe et al., 2000). CV–N binds to recombinant soluble gp120 with a K_d of 2–45 nM and a 5:1 stoichiometry; and to soluble gp41 with K_d of 606 nM and 1:1 stoichiometry (O'Keefe et al., 2000).

The antiviral activity of CV–N is mediated through specific, high-affinity interactions with the viral surface envelope glycoproteins, gp120 (Boyd et al., 1997) and possibly also gp41 (O'Keefe, 2001; O'Keefe et al., 2000). It is thought that binding of CV–N renders these glycoproteins incapable of mediating virus-to-cell or cell-to-cell fusion.

A number of structures of wild-type, mutant, and sequence-shuffled CV–N have been solved by NMR (Bewley et al., 1998; Clore and Gronenborn, 1998) and crystallography (Barrientos et al., 2002a; Botos et al., 2002a, b; Yang et al., 1999), showing that the protein exists as either a quasi-symmetric two-domain monomer or a domain-swapped dimer (see Table 7). Depending on the pH of the crystallization condition, two crystal forms were observed: trigonal (space group $P3_221$) grown at low pH (Yang et al., 1999), and tetragonal (space group $P4_12_12$) grown at high pH (Barrientos et al., 2002a). Some of the structural features appear to be pH-dependent.

As expected from the sequence similarity of parts A and B of CV–N, the solution structure of monomeric protein revealed two very similar domains (Bewley et al., 1998). However, the first of these domains was made up of residues 1–38 and 90–101, and the second one from residues 39–89. Since the structure of the hairpin 90–101 is almost identical to that of 39–50, the monomer has been formed by intramolecular strand exchange, usually called domain swapping (Bennett et al.,

Table 7
Currently available structures of cyanovirin

PDB ID	Method, reference	Resolution (Å)	Ligand	CV–N	Oligomeric form	Space group
2EZN	NMR (Bewley et al., 1998)	—	—	Wild-type	Compact monomer	—
2EZM	NMR (Bewley et al., 1998)		—	Wild-type	Compact monomer	—
3EZM	X-ray (Yang et al., 1999)	1.5	—	Wild-type	Domain-swap dimer	$P3_221$
1J4V	NMR (Clore and Bewley, 2002)	—	—	Wild-type	Domain-swap dimer	—
1L5B	X-ray (Barrientos et al., 2002a)	2.0	—	Wild-type	Domain-swap dimer	$P4_12_12$
1L5E	NMR (Barrientos et al., 2002a)	—	—	Wild-type	Domain-swap dimer	—
1IIY	NMR (Bewley, 2001b)	—	α-Man-(1→2)-α-Man	Wild-type	Compact monomer	—
1M5M	X-ray (Botos et al., 2002b)	2.5	Man-9	Wild-type	Domain-swap dimer	$P4_12_12$
1M5J	X-ray (Botos et al., 2002b)	2.4	Hexamannose	Wild-type	Domain-swap dimer	$P4_12_12$
1LOM	X-ray (Botos et al., 2002a)	1.72	—	P51S/S52P mutant	Domain-swap dimer	$P3_221$
1NO2	NMR (Barrientos et al., 2003)	—	—	Circularly permuted (cpCV–N)	Compact monomer	—
—	NMR (Bewley, 2001a)	—	—	sequence-swapped (dsCV–N)	Compact monomer	—

1995). The crystal structure of the CV–N dimer (Yang et al., 1999) represents an unusual case of double domain swapping and consists of two quasi-monomers resembling the one described above, with one made up of residues 1–50 (A) and 51′–101′ (B′) and the other of 51–101 (B) and 1′–50′ (A′). Similar dimers were later also observed in some NMR structures (Barrientos et al., 2002a; Bewley and Clore, 2000; Clore and Bewley, 2002).

A major difference between the initial solution and crystal structures of CV–N was the three-dimensional domain swapping, observed in the latter case. Due to a change of torsion angles in the hinge region (residues 49–54), domains A and B of CV–N separate into a very elongated entity in which they do not contact each other (see Fig. 8C). Two such symmetrically related monomers form a dimer in which domain A from one monomer interacts with domain B′ from the other monomer and vice versa, forming two quasi-monomers, AB′ and A′B, linked by their respective hinge regions (see Fig. 8C). This domain-swapped dimer is quite unique in that half of each molecule takes part in its creation (in most other cases described to date, only a small part of the protein is involved in domain swapping (Bennett et al., 1995)). The quasi-monomers can adopt a range of orientations relative to each other, depending on the pH of the environment.

The reasons behind these large differences in the orientation of the domains have not yet been satisfactorily explained. It has been shown that proline residues are able to influence protein

conformation (MacArthur and Thornton, 1991) and, when located in a hinge region, may promote arm exchange and domain swapping for oligomerization of monomeric proteins (Rousseau et al., 2001). However, the importance of these proline residues was put into a different light by the crystal structure of a CV–N double-mutant Pro51Ser/Ser52Pro (Botos et al., 2002a). In the Pro51Ser/Ser52Pro structure the hinge region adopts a significantly different orientation than in the wild-type domain-swapped dimer, although the relative orientation of quasi-monomers is the same as in the low-pH crystals of wild-type CV–N. However, this mutant adopts the same relative orientation of the quasi-monomers under a wide range of pH conditions, challenging the hypothesis that domain-swapping is a result of low pH and further questioning the relative importance of the proline residue in the hinge region as a determinant of domain orientation.

A surface and volume analysis of some of the available CV–N structures (Botos et al., 2002a) showed a trend of more buried atoms with the decrease of pH, yielding a more compact structure at low pH. At the same time, the solution structures proved to be more compact than the crystal structures. These observations suggest a domain orientation determined by electrostatics of the whole protein rather than a single proline residue in the hinge region. The relative orientation of the domains may be important, since it might influence the oligosaccharide-binding affinity of CV–N. The dimer of CV–N was shown to be a metastable, kinetically trapped intermediate at neutral pH and room temperature, under which conditions it is stable for months. After incubation for 24 h at 38 °C, the metastable dimer converts into a thermodynamically more stable monomer (Barrientos et al., 2002a).

Two carbohydrate-binding sites with different affinities have been mapped on the surface of CV–N (Bewley and Otero-Quintero, 2001). These sites are created by structurally equivalent residues belonging to the two domains of monomeric CV–N and bind primarily N-linked high-mannose oligosaccharides, such as those found on the viral envelope of HIV-1 (Bolmstedt et al., 2001). In monomeric CV–N, the primary carbohydrate-binding site is defined by residues 41–44, 50–56, and 74–78, and the secondary site by residues 1–7, 22–26 and 92–95 (Bewley and Otero-Quintero, 2001). Four sugar-binding sites can be identified in a domain-swapped dimer. The two primary sites are located close to each other near the hinge region, and two secondary sites are found on the opposite sides of the dimer, thus they are not influenced directly by the conformation of the hinge region. In a monomer, intertwined loops from domains A and B form the sugar-binding sites, whereas in a domain-swapped dimer these loops contain residues from both monomers. In this manner, residues 1–7 and 22–26 from the first monomer and 92′–95′ and 101′ from the second monomer form the secondary site.

The structure of the primary sugar-binding site of CV–N was first described in solution, in the presence of the disaccharide α-Man-(1→2)-α-Man (Bewley, 2001b). The disaccharide binds in a deep pocket in the close proximity of the hinge region in a stacked conformation, establishing hydrogen bonds with Glu41, Asp44, Ser52, Glu56, Thr57, Thr75, Arg76, and Gln78 (see Fig. 8D). In the crystal structure of a complex of CV–N with an oligosaccharide (see Figs. 8E and F), the primary sugar-binding site contained, instead of a sugar, a well-defined, tightly bound buffer molecule (CHES) from the crystallization solution (Botos et al., 2002b).

The variable geometry of the primary sugar-binding site upon domain swapping was revealed by crystal (Barrientos et al., 2002a; Yang et al., 1999) and solution structures (Bewley, 2001b; Bewley et al., 1998) of CV–N. Since the primary sugar-binding site is in close proximity to the hinge region, the position of the hinge and relative orientation of the domains has a direct impact

on the shape of the primary sugar-binding site. In the monomer, the pocket is intact (see Fig. 8G), accommodating a disaccharide in a stacked conformation (Bewley, 2001b), whereas in all domain-swapped structures some of the essential protein–oligosaccharide hydrogen bonds cannot be established, leaving only the shape complementarity of oligosaccharide and pocket. Specifically, in the low pH (trigonal) crystal structure, the sugar-binding pocket is partially changed compared to the one observed in the compact monomeric structure (see Fig. 8(H)). Displacement of the Ser52 OG atom disrupts its potential hydrogen-bonding to a mannose ring, reducing the binding affinity of the site. This is the case in all trigonal structures crystallized at pH range 4.6–8.5, which adopt the same domain orientation and do not bind oligosaccharide in the primary site (Botos et al., 2002b). High pH (tetragonal) crystal structures have a different relative orientation of the domains but the side chain of Asn53 from the hinge moves into the pocket, perturbing its geometry (see Fig. 8I). Because of these steric considerations, sugar binding in this pocket may not be favorable and was not observed in any of the tetragonal crystal structures at various pH values. Since binding into the primary sugar-binding site was not observed, the question arises whether the binding into this pocket is also carried out by three stacked rings like in the secondary binding site, or by two stacked rings only, as previously reported (Bewley, 2001b).

A secondary sugar-binding site is located ~35 Å from the primary site on the same monomer, or quasi-monomer. This site has a binding interface formed by three $\alpha 1 \rightarrow 2$ linked, stacked rings in the case of Man-9 (see Fig. 8E) and two in the case of hexamannose (see Fig. 8F) (Botos et al., 2002b). The position of mannose rings C and 4 from arm D1 is very similar to the model suggested from the solution structure of a CV–N disaccharide complex (Bewley, 2001b). However, the stacked rings interact much tighter with the protein than is suggested by the NMR model, with nine hydrogen bonds between the Man-9 rings and CV–N. Since the only difference between the D1 branches in Man-9 and hexamannose is the lack of one $\alpha 1 \rightarrow 2$ linked mannose ring at the end of the branch in the latter, the crystal structures unequivocally identify the D1 arm as the binding interface. This binding site, located far from the hinge region, is not affected by the geometry of the hinge, and presents the same conformation in both monomeric and domain-swapped dimeric CV–N (Botos et al., 2002b).

Comparison of different structures shows that they differ mainly around their carbohydrate binding sites (Botos et al., 2002a). CD measurements show that conformational changes occur upon CV–N binding to gp120 and gp41 (Esser et al., 1999; Mariner et al., 1998; O'Keefe et al., 2000), resulting in an average 11% loss of β-sheet and 2% loss of helical structure. Analysis of temperature factors of the oligosaccharide-binding site in the uncomplexed and sugar-bound forms of CV–N shows higher than average B-factors for the sugar-binding site residues.

The anti-HIV activity and gp120 binding have been determined for wild-type CV–N, and a series of engineered mutants and other constructs (Mori et al., 1997a) (see Table 8). The sequence specificity for anti-HIV activity and gp120 binding do not appear to be identical. The key structural regions responsible for activity were established by some of the early constructs (Mori et al., 1997a). However, structural studies of CV–N opened the possibility of rationally designing specific mutants, which could drastically alter some of the characteristics of the protein. Mutations in the hinge region can influence the dynamics of the two domains, preventing three-dimensional domain swapping or locking the protein in an extended conformation.

Ser52Pro is a hinge-region mutant of CV–N that exists in solution exclusively as a stable dimer (Han et al., 2002) and shows lower potency against HIV compared to the wild-type CV–N. The

Table 8
Summary of biological activity of CV–N and its mutants

Protein	Order of residues	EC_{50} (nM)
Wild type CVN[1]	1–101	0.9 ± 0.4
CVN–1[1]	2–101	2.0 ± 0.7
CVN–2[1]	3–101	8.3 ± 4.0
CVN–3[1]	4–101	140.7 ± 36.7
F–CVN[1]	1–101	3.6 ± 1.4
F–CVN[1]	1–98	149.5 ± 17.2
F–CVN[1]	1–93	Inactive
F–D1D1[1]	1–50, 1–50	Inactive
F–D2D2[1]	52–101, 51–101	Inactive
F–D2D1[1]	52–101, 1–50	217.7 ± 9.5
F–C8S–C22S[1]	1–101	Inactive
F–C58S–C73S[1]	1–101	Inactive
		EC_{3D} (nM)
Wild type CV–N[2]	1–101	0.19
A77T[2]	1–101	0.22
S52P[2]	1–101	1.46
cpCVN[3]	1–3, 54–98, 48–53, 4–47, 99–101	1000 × less potent than wt
dsCVN[4]	1–3, 54–98, 48–53, 4–47, 99–101	500 × less potent than wt

F indicates uncleaved FLAG-tag used in protein expression (Mori et al., 1997a). The second column shows the order of residues in the various constructs. The XTT-tetrazolium assay[1,2,3] (Gulakowski et al., 1991), and an HIV-1 envelope-mediated cell fusion assay[4] were used for biological activity determination.Compiled from the following references: Mori et al. (1997a)[1], Han et al. (2002)[2], Barrientos et al. (2002b)[3]and Bewley (2001a)[4]. Since activity values were obtained using different assays, they are not directly comparable.

Pro51Gly mutant is mainly monomeric in solution with only trace amounts of dimers and is more stable than wild-type CV–N under certain experimental conditions (Mori et al., 1998). A double mutant in these two positions, Pro51Ser/Ser52Pro, exhibits equivalent properties and similar structure to wild-type CV–N (Botos et al., 2002a). However, the mutant shows a different orientation in the hinge region that connects two domains of the protein.

An obligate domain-swapped mutant generated by deletion of Gln50 from the hinge region is dimeric and domain-swapped regardless of pH, with slightly more potent HIV-1 fusion inhibitory activity than wild-type CV–N (Kelley et al., 2002). The relevance of the latter result to the more pertinent question of actual antiviral activity against clinical HIV isolates is unknown.

Carbohydrate binding in the secondary site was completely abolished in a triple mutant Lys3Asn, Glu23Ile, Asn93Ala, and in a quadruple mutant Lys3Asn, Thr7Asn, Glu23Ile, Asn93Ala, while α-Man-(1→2)-α-Man was bound in the primary site as strongly as in the wild-type CV–N (Chang and Bewley, 2002). All of the mutants inhibited HIV-1 fusion nearly identically to wild-type CV–N. These results were interpreted as indicating that the secondary binding site is not necessary for high affinity binding to gp120, with monovalent protein–oligosaccharide interactions being sufficient for blocking HIV-1 fusion (Chang and Bewley, 2002).

dsCV–N is an internal sequence-swapped mutant with the cores of domains A and B swapped, with three terminal amino acids and the linker unaltered (Bewley, 2001a). In a quantitative HIV-1

envelope-mediated cell fusion assay (Nussbaum et al., 1994), dsCV–N was 500-fold less active than wild-type CV–N. dsCV–N has the same overall shape as wild-type CV–N (Bewley, 2001a) and can also be described as a variation of a circularly permuted form of CV–N (cpCV–N) in which several residues are shifted in their sequence (Barrientos et al., 2001, 2002b). cpCV–N is monomeric and less stable than wild-type CV–N, with significantly reduced anti-HIV activity (Barrientos et al., 2003).

Titration calorimetric binding experiments of CV–N to Man-8 show enthalpically driven binding (Shenoy et al., 2001). It was shown that protein lectins and oligosaccharides can overcome weak individual interaction free energies of binding by mediating multivalent binding with one another (Dimick et al., 2002; Rice and Lee, 1993). In addition to the protein–oligosaccharide interactions, specific protein–protein interactions may possibly play an important role in enhancing the CV–N gp120 binding, since sgp120 mediates a 13-fold greater affinity with CV–N than Man-8 (Shenoy et al., 2001). The following CV–N/gp120 binding model was proposed (Bewley and Otero-Quintero, 2001). It is likely that CV–N binds to high-mannose saccharides on gp120 via the high affinity oligosaccharide-binding site. Binding creates a high apparent oligomannose concentration and CV–N could further interact with an appropriately spaced second oligomannose via the lower-affinity site. Binding to oligomannose on gp120 through both oligosaccharide-binding sites will result in multivalent binding that can facilitate crosslinking, observed with Man-8 and Man-9 at micromolar to millimolar concentrations (Bewley and Otero-Quintero, 2001). It is possible that local pH conditions regulate gp120 binding by influencing the quasi-monomer orientation in the domain-swapped CV–N.

The binding specificity of CV–N to gp120 might be explained by its ability to distinguish between high-mannose and complex oligosaccharides (Shenoy et al., 2001). The binding sites for CV–N might overlap in regions containing high-mannose (Esser et al., 1999).

CV–N preferentially inhibits binding of the glycosylation-dependent neutralizing monoclonal antibody 2G12–gp120 (Boyd et al., 1997; Esser et al., 1999), and also interferes with the binding of cell-associated CD4 to virion-associated gp120, but not of the soluble CD4 (sCD4) with either soluble gp120 (sgp120) or virion-associated gp120 (Mori and Boyd, 2001). CV–N impairs both CD4-dependent and CD4-independent binding of gp120 to the target cells, blocks the sCD4-induced binding of sgp120 with cell-associated coreceptor CXCR4 and dissociates bound sgp120 from target cells (Mori and Boyd, 2001).

The physiological role of CV–N within cyanobacteria and the significance, if any, of CV–N inhibiting retroviruses are presently unknown (Gustafson et al., 1997). Due to its broad spectrum of activity against immunodeficiency viruses, CV–N is promising in both prevention and treatment strategies for AIDS.

4.13. Scytovirin

Scytovirin is a recently discovered anti-HIV protein, isolated from aqueous extracts of the cultured cyanobacterium *Scytonema varium* (Bokesch et al., 2003). Scytovirin is a 95 amino acid protein with a molecular mass of 9713 Da, with five intrachain disulfide bonds. The disulfide bond pattern was analyzed by a trypsin digest followed by mass spectrometry of the fragments. The five disulfide bonds are formed between Cys20–Cys26, Cys32–Cys38, Cys68–Cys74, Cys80–Cys86 and Cys7–Cys55 (Bokesch et al., 2003).

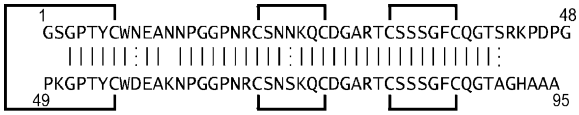

Fig. 9. Internal duplication in the amino acid sequence of scytovirin. Disulfide bonds are indicated by solid lines.

The amino acid sequence reveals the presence of strong internal sequence duplication (see Fig. 9). When residues 1–48 are aligned to residues 49–95, 36 of them are identical (75%) and 6 are homologous (12%). Disulfide bonds Cys20–Cys26 and Cys32–Cys38 correspond to the bonds defined by Cys68–Cys74 and Cys80–Cys86. This pattern suggests two functional domains linked by Cys7–Cys55 (Bokesch et al., 2003).

The amino acid sequence of a domain of scytovirin is 55% homologous to the chitin-binding domain from the green algae *Volvox carteri*. The chitin-binding domain consists of a 30–43 residue structural motif with glycines and cysteines at conserved positions, however, with a different disulfide-bonding pattern than that of scytovirin. Scytovirin has lower sequence homology with precursor proteins of *Urtica dioica* agglutinin isolectins, the conidiospore surface protein and the thermal hysteresis proteins (THPs). The six-residue cysteine spacing in THPs is thought to have a structural role and contribute to the antifreeze function. Scytovirin does not seem to belong to the chitin-binding class of proteins, since it does not bind chitin and has a distinct disulfide-bonding pattern (Bokesch et al., 2003).

Scytovirin has strong activity against T-tropic laboratory-adapted strains and primary isolates of HIV-1, but is 300-fold less effective against M-tropic strains. For antiviral activity, scytovirin has to be present within the first 8 h of viral infection. Removal of scytovirin caused normal susceptibility to HIV infection and normal infectivity in cell-free virus (Bokesch et al., 2003). The viral inhibition by scytovirin is reversible, in contrasts with the mode of action of CV–N which potently inactivates the virus on contact (Boyd et al., 1997; Gustafson et al., 1997).

ELISA studies indicated that scytovirin binds specifically to glycoproteins gp120 and gp160, and to a lesser extent to gp41. When tested for monosaccharide binding, unlike some other lectins, scytovirin does not show specificity for D-Man or GlcNAc. However, scytovirin binding to gp120 can be inhibited by oligomannose 8 and oligomannose 9, suggesting that the protein might interact preferentially with high-mannose sites on the surface of gp120 (Bokesch et al., 2003). These studies show that scytovirin is an inhibitor of HIV binding and/or fusion, but its mode of binding of oligosaccharides is not known due to the lack of structural information on this protein.

5. Concluding remarks

Many proteins that bind HIV-1 gp120 envelope carbohydrates have already been characterized. The intimate mode of interaction of these proteins with high-mannose carbohydrates was reviewed here principally from a structural point of view. Carbohydrates are relatively polar molecules, but their interactions with proteins are not solely dominated by hydrogen bonding. Besides direct and water-mediated hydrogen bond networks, van der Waals and hydrophobic interactions also make a major contribution to the stability of the carbohydrate–protein

complexes. Aromatic residues in key positions pack against sugar rings, or methyl groups from *N*-acetyl moieties (Drickamer, 1995).

Although the modes of interaction of proteins with the carbohydrates are strikingly different, some common motifs of binding can be distinguished. The first of these motifs relates to the role of a metal ion in sugar binding. Typical C-type lectins bind carbohydrates directly via a Ca^{2+} atom. While the other proteins do not involve a metal ion in direct carbohydrate binding, ConA uses metal ions indirectly to stabilize the sugar-binding site.

Another motif is the specificity for a particular linkage between carbohydrate rings. A group of proteins (MBPs, MHL, UDA, monocot lectins, CV–N, 2G12) interacts with terminal carbohydrates (Man, GlcNAc), while others prefer interactions with oligosaccharide cores (DC-SIGN, DC-SIGNR, ConA, monocot lectins). MBPs interact with a single terminal carbohydrate ring, while other proteins from the above group can interact with several carbohydrate rings of the oligosaccharide. Some proteins are specific for $\alpha 1 \rightarrow 2$ linked Man residues (CV–N, 2G12), or $\alpha 1 \rightarrow 3$ and $\alpha 1 \rightarrow 6$ linked residues (monocot lectins, ConA, DC-SIGN, DC-SIGNR), while others do not show high specificity.

Finally, there are some lectins for which the interaction mode has not yet been elucidated. Future structural studies of these proteins will permit a better understanding of carbohydrate–protein interactions.

Monosaccharide–lectin interactions are, in general, relatively weak, with binding affinities in the millimolar range. Some binding sites show multiple specificities, like that of MPBs for Man, GlcNac, and Fuc. The orientation of a few hydroxyl groups is the only common feature of these carbohydrates. Slight modifications of the sugar binding site can dramatically change the binding selectivity (Drickamer, 1995).

One might wonder, where does the high selectivity of these proteins come from? The review of the above structures shows an abundance of extended, secondary sugar binding sites. Several sugar residues of an oligosaccharide can interact with such an extended binding site. Another source for higher affinity is the multiplicity of binding sites. In the case of monocot family lectins there are 3–4 binding sites per molecule, which amount to 12–16 sites per complete lectin tetramer. Such an abundance of binding sites can compensate for the weak affinity of individual sites. Multiple binding sites can bind terminal sugars of individual oligosaccharides or terminal sugars of the same multiantennary oligosaccharide molecule (subunit multivalency). These proteins may have a bridging function, linking different cells via their oligosaccharides (Drickamer, 1995; Weis and Drickamer, 1996).

While the high density of oligosaccharides on gp120 is thought to play a protective role by reducing the immunogenicity of gp120, it may be possible to use these surface carbohydrates as targets in novel antiviral strategies (Hart et al., 2003). Alteration of virus carbohydrates in vivo could increase the amount of MBL-binding carbohydrates on the virus surface, resulting in more efficient clearance and/or neutralization of the virus. Both the treatment with neuraminidase and glycosylation inhibitors strengthen the interaction of HIV-1 with MBL, enhancing MBL-mediated neutralization of HIV-1 (Hart et al., 2003).

HIV has a high mutation rate, rapidly evolving drug-resistant strains after administration of most anti-HIV drugs, further complicating development of vaccines (Burton, 1997; Heilman and Baltimore, 1998). A drug that could effectively prevent initial viral infection would be more likely

to prevent the emergence of a drug-resistant virus. There is a growing interest in development of anti-HIV microbicides, that may be used either topically or ex vivo to prevent the spread of viral infection (Lange et al., 1993), CV–N being one of the prime candidates (Boyd et al., 1997; Dey et al., 2000; Esser et al., 1999).

CV–N can also be used for creating a recombinant chimeric toxin molecule in which the translocation and cytotoxic domains of *Pseudomonas* exotoxin A are linked to CV–N. In the resulting molecule, CV–N serves as the targeting moiety that binds to HIV-infected cells expressing gp120. The chimeric molecule shows enhanced cytotoxicity for HIV-infected H9 cells compared to uninfected H9 cells, although with diminution of the overall gp120-binding activity (Mori et al., 1997b).

CV–N is possibly immunogenic in humans, therefore its use for ex vivo inactivation of HIV in blood or plasma was explored (Gandhi et al., 2000). Biotinylated CV–N can be coupled to streptavidin coated magnetic beads creating sessile CV–N. Sessile CV–N completely removed 100 $TCID^{50}$ of the infectious virions, leaving behind replication-incompetent virions that could be used as a vaccine for HIV.

CV–N is currently in preclinical development for use as a topical microbicide to prevent genital HIV transmission (De Clercq, 2000). An effective CV–N-based microbicide may require long-term presence or continuous delivery of CV–N at the mucosal sites of entry of HIV. A possible approach is endogenous production of virucidal concentrations of CV–N at mucosal surfaces by a colonizing commensal bacterium such as *Streptococcus gordonii* (Giomarelli et al., 2002). Work is in progress exploring the possibility of expression of CV–N in *Lactobacillus* strains (Giomarelli et al., 2002). Other proteins described in this review may ultimately also be useful in clinical applications against retroviral infection.

Acknowledgements

The authors would like to thank Dr. Jeannine Botos for editorial suggestions.

References

Allan, J.S., Coligan, J.E., Barin, F., McLane, M.F., Sodroski, J.G., Rosen, C.A., Haseltine, W.A., Lee, T.H., Essex, M., 1985. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. Science 228, 1091–1094.

Aucouturier, P., Mihaesco, E., Mihaesco, C., Preud'Homme, J.L., 1987. Characterization of jacalin, the human IgA and IgD binding lectin from jackfruit. Mol. Immunol. 24, 503–511.

Balzarini, J., Schols, D., Neyts, J., Van Damme, E., Peumans, W., De Clercq, E., 1991. Alpha-(1-3)- and alpha-(1-6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections in vitro. Antimicrob. Agents Chemother. 35, 410–416.

Balzarini, J., Neyts, J., Schols, D., Hosoya, M., Van Damme, E., Peumans, W., De Clercq, E., 1992. The mannose-specific plant lectins from *Cymbidium hybrid* and *Epipactis helleborine* and the (*N*-acetylglucosamine)*n*-specific plant lectin from *Urtica dioica* are potent and selective inhibitors of human immunodeficiency virus and cytomegalovirus replication in vitro. Antiviral Res. 18, 191–207.

Banchereau, J., Steinman, R.M., 1998. Dendritic cells and the control of immunity. Nature 392, 245–252.

Barrientos, L.G., Campos-Olivas, R., Louis, J.M., Fiser, A., Sali, A., Gronenborn, A.M., 2001. ^{1}H, ^{13}C, ^{15}N resonance assignments and fold verification of a circular permuted variant of the potent HIV-inactivating protein cyanovirin–N. J. Biomol. NMR 19, 289–290.

Barrientos, L., Louis, J.M., Botos, I., Mori, T., Han, Z., O'Keefe, B.R., Boyd, M.R., Wlodawer, A., Gronenborn, A., 2002a. The domain-swapped dimer of Cyanovirin–N is in a metastable folded state. Reconcilliation of X-ray and NMR structures. Structure 10, 673–686.

Barrientos, L.G., Louis, J.M., Hung, J., Smith, T.H., O'Keefe, B.R., Gardella, R.S., Mori, T., Boyd, M.R., Gronenborn, A.M., 2002b. Design and initial characterization of a circular permuted variant of the potent HIV-inactivating protein Cyanovirin–N. Proteins 46, 153–160.

Barrientos, L.G., Louis, J.M., Ratner, D.M., Seeberger, P.H., Gronenborn, A.M., 2003. Solution structure of a circular-permuted variant of the potent HIV-inactivating protein Cyanovirin–N: structural basis for protein stability and oligosaccharide interaction. J. Mol. Biol. 325, 211–223.

Bashirova, A.A., Geijtenbeek, T.B., van Duijnhoven, G.C., van Vliet, S.J., Eilering, J.B., Martin, M.P., Wu, L., Martin, T.D., Viebig, N., Knolle, P.A., KewalRamani, V.N., van Kooyk, Y., Carrington, M., 2001. A dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin (DC-SIGN)-related protein is highly expressed on human liver sinusoidal endothelial cells and promotes HIV-1 infection. J. Exp. Med. 193, 671–678.

Beintema, J.J., Peumans, W.J., 1992. The primary structure of stinging nettle (*Urtica dioica*) agglutinin. A two-domain member of the hevein family. FEBS Lett. 299, 131–134.

Bennett, M.J., Schlunegger, M.P., Eisenberg, D., 1995. 3D domain swapping: a mechanism for oligomer assembly. Protein Sci. 4, 2455–2468.

Berger, E.A., Murphy, P.M., Farber, J.M., 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu. Rev. Immunol. 17, 657–700.

Bergsma, J., Hol, W.G., Jansonius, J.N., Kalk, K.H., Ploegman, J.H., Smit, J.D., 1975. The double domain structure of rhodanese. J. Mol. Biol. 98, 637–643.

Bewley, C.A., 2001a. Rapid validation of the overall structure of an internal domain-swapped mutant of the anti-HIV protein Cyanovirin–N using residual dipolar couplings. J. Am. Chem. Soc. 123, 1014–1015.

Bewley, C.A., 2001b. Solution structure of a Cyanovirin–N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120. Structure 9, 931–940.

Bewley, C.A., Clore, G.M., 2000. Determination of the relative orientation of the two halves of the domain-swaped dimer of Cyanovirin–N in solution using dipolar couplings and rigid body minimization. J. Am. Chem. Soc. 122, 6009–6016.

Bewley, C.A., Otero-Quintero, S., 2001. The potent anti-HIV protein Cyanovirin–N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. J. Am. Chem. Soc. 123, 3892–3902.

Bewley, C.A., Gustafson, K.R., Boyd, M.R., Covell, D.G., Bax, A., Clore, G.M., Gronenborn, A.M., 1998. Solution structure of Cyanovirin–N, a potent HIV-inactivating protein. Nature Struct. Biol. 5, 571–578.

Bokesch, H.R., O'Keefe, B.R., McKee, T.C., Pannell, L.K., Patterson, G.M., Gardella, R.S., Sowder, R.C., Turpin, J., Watson, K., Buckheit Jr., R.W., Boyd, M.R., 2003. A potent novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. Biochemistry 42, 2578–2584.

Bolmstedt, A.J., O'Keefe, B.R., Shenoy, S.R., McMahon, J.B., Boyd, M.R., 2001. Cyanovirin–N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. Mol. Pharmacol. 59, 949–954.

Botos, I., Wlodawer, A., 2003. Cyanovirin-N: a sugar-binding antiviral protein with a new twist. Cell. Mol. Life Sci. 60, 277–287.

Botos, I., Mori, T., Cartner, L.K., Boyd, M.R., Wlodawer, A., 2002a. Domain-swapped structure of a mutant of Cyanovirin–N. Biochem. Biophys. Res. Commun. 294, 184–190.

Botos, I., O'Keefe, B.R., Shenoy, S.R., Cartner, L.K., Ratner, D.M., Seeberger, P.H., Boyd, M.R., Wlodawer, A., 2002b. Structures of the complexes of a potent anti-HIV protein Cyanovirin–N and high mannose oligosaccharides. J. Biol. Chem. 277, 34336–34342.

Bouckaert, J., Loris, R., Poortmans, F., Wyns, L., 1995. Crystallographic structure of metal-free concanavalin A at 2.5 Å resolution. Proteins 23, 510–524.

Bouckaert, J., Poortmans, F., Wyns, L., Loris, R., 1996. Sequential structural changes upon zinc and calcium binding to metal-free concanavalin A. J. Biol. Chem. 271, 16144–16150.

Bouckaert, J., Hamelryck, T.W., Wyns, L., Loris, R., 1999. The crystal structures of Man(alpha1-3)Man(alpha1-O)Me and Man(alpha1-6)Man(alpha1-O)Me in complex with concanavalin A. J. Biol. Chem. 274, 29188–29195.

Bourne, Y., Astoul, C.H., Zamboni, V., Peumans, W.J., Menu-Bouaouiche, L., Van Damme, E.J., Barre, A., Rouge, P., 2002. Structural basis for the unusual carbohydrate-binding specificity of jacalin towards galactose and mannose. Biochem. J. 364, 173–180.

Boyd, M.R., Gustafson, K.R., McMahon, J.B., Shoemaker, R.H., O'Keefe, B.R., Mori, T., Gulakowski, R.J., Wu, L., Rivera, M.I., Laurencot, C.M., Currens, M.J., Cardellina, J.H., Buckheit Jr., R.W., Nara, P.L., Pannell, L.K., Sowder, R.C., Henderson, L.E., 1997. Discovery of Cyanovirin–N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. Antimicrob. Agents Chemother. 41, 1521–1530.

Bradbrook, G.M., Gleichmann, T., Harrop, S.J., Habash, J., Raftery, J., Kalb, J.G., Yariv, J., Hillier, I.H., Helliwell, J.R., 1998. X-ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes. J. Chem. Soc., Faraday Trans. 94, 1603–1611.

Brewer, C.F., Brown III, R.D., Koenig, S.H., 1983. Stoichiometry of manganese and calcium ion binding to concanavalin A. Biochemistry 22, 3691–3702.

Burton, D.R., 1997. A vaccine for HIV type 1: the antibody perspective. Proc. Natl. Acad. Sci. USA 94, 10018–10023.

Calarese, D.A., Scanlan, C.N., Zwick, M.B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M.R., Stanfield, R.L., Roux, K.H., Kelly, J.W., Rudd, P.M., Dwek, R.A., Katinger, H., Burton, D.R., Wilson, I.A., 2003. Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. Science 300, 2065–2071.

Chan, D.C., Kim, P.S., 1998. HIV entry and its inhibition. Cell 93, 681–684.

Chan, D.C., Fass, D., Berger, J.M., Kim, P.S., 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell 89, 263–273.

Chang, L.C., Bewley, C.A., 2002. Potent inhibition of HIV-1 fusion by Cyanovirin–N requires only a single high affinity carbohydrate binding site: characterization of low-affinity carbohydrate binding site knockout mutants. J. Mol. Biol. 318, 1–8.

Chantalat, L., Wood, S.D., Rizkallah, P., Reynolds, C.D., 1996. X-ray structure solution of amaryllis lectin by molecular replacement with only 4% of the total diffracting matter. Acta Crystallogr. D 52, 1146–1152.

Chapot, M.-P., Peumans, W.J., Strosberg, A.D., 1986. Extensive homologies between lectins from non-leguminous plants. FEBS Lett. 195, 231–234.

Charan, R.D., Munro, M.H., O'Keefe, B.R., Sowder, R.C.I.I., McKee, T.C., Currens, M.J., Pannell, L.K., Boyd, M.R., 2000. Isolation and characterization of *Myrianthus holstii* lectin, a potent HIV-1 inhibitory protein from the plant *Myrianthus holstii* (1). J. Nat. Prod. 63, 1170–1174.

Childs, R.A., Feizi, T., Yuen, C.T., Drickamer, K., Quesenberry, M.S., 1990. Differential recognition of core and terminal portions of oligosaccharide ligands by carbohydrate-recognition domains of two mannose-binding proteins. J. Biol. Chem. 265, 20770–20777.

Christopher, J.A., 1998. SPOCK: The Structural Properties Observation and Calculation Kit. The Center for Macromolecular Design, Texas A&M University, College Station, TX.

Clore, G.M., Bewley, C.A., 2002. Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. J. Magn. Reson. 154, 329–335.

Clore, G.M., Gronenborn, A.M., 1998. New methods of structure refinement for macromolecular structure determination by NMR. Proc. Natl. Acad. Sci. USA 95, 5891–5898.

Coffin, J.M., Huges, S.H., Varmus, H.E., 1997. Retroviruses. Cold Spring Harbor Laboratory Press, New York.

Corbeau, P., Haran, M., Binz, H., Devaux, C., 1994. Jacalin, a lectin with anti-HIV-1 properties, and HIV-1 gp120 envelope protein interact with distinct regions of the CD4 molecule. Mol. Immunol. 31, 569–575.

Corbeau, P., Pasquali, J.L., Devaux, C., 1995. Jacalin, a lectin interacting with O-linked sugars and mediating protection of CD4+ cells against HIV-1, binds to the external envelope glycoprotein gp120. Immunol. Lett. 47, 141–143.

De Clercq, E., 2000. Current lead natural products for the chemotherapy of human immunodeficiency virus (HIV) infection. Med. Res. Rev. 20, 323–349.

Deacon, A., Gleichman, T., Kalb (Gilboa), A.J., Price, H., Raftery, J., Bradbrook, G., Yariv, J., Helliwell, J.R., 1997. The structure of concanavalin A and its bound solvent determined with small-molecule accuracy at 0.94 Å resolution. J. Chem. Soc., Faraday Trans. 93, 4305–4312.

Dey, B., Lerner, D.L., Lusso, P., Boyd, M.R., Elder, J.H., Berger, E.A., 2000. Multiple antiviral activities of Cyanovirin–N: blocking of human immunodeficiency virus type 1 gp120 interaction with CD4 and co-receptor and inhibition of diverse enveloped viruses. J. Virol. 74, 4562–4569.

Dimick, S.M., Powel, S.C., McMahon, S.A., Moothoo, D.N., Haismith, J.H., Toone, E.J., 2002. On the meaning of affinity: cluster glycoside effects and concanavalin A. J. Am. Chem. Soc. 121, 10286–10296.

Doms, R.W., Peiper, S.C., 1997. Unwelcomed guests with master keys: how HIV uses chemokine receptors for cellular entry. Virology 235, 179–190.

Doranz, B.J., Filion, L.G., Diaz-Mitoma, F., Sitar, D.S., Sahai, J., Baribaud, F., Orsini, M.J., Benovic, J.L., Cameron, W., Doms, R.W., 2001. Safe use of the CXCR4 inhibitor ALX40-4C in humans. AIDS Res. Hum. Retroviruses 17, 475–486.

Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P., Paxton, W.A., 1996. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 381, 667–673.

Drickamer, K., 1995. Multiplicity of lectin–carbohydrate interactions. Nature Struct. Biol. 2, 437–439.

Drickamer, K., 1999. C-type lectin-like domains. Curr. Opin. Struct. Biol. 9, 585–590.

Drickamer, K., Taylor, M.E., 1993. Biology of animal lectins. Annu. Rev. Cell Biol. 9, 237–264.

Earl, P.L., Moss, B., 1993. Mutational analysis of the assembly domain of the HIV-1 envelope glycoprotein. AIDS Res. Hum. Retroviruses 9, 589–594.

Earl, P.L., Doms, R.W., Moss, B., 1990. Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. Proc. Natl. Acad. Sci. USA 87, 648–652.

Earl, P.L., Moss, B., Doms, R.W., 1991. Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein. J. Virol. 65, 2047–2055.

Esnouf, R.M., 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. J. Mol. Graph. Model. 15, 132–134.

Esser, M.T., Mori, T., Mondor, I., Sattentau, Q.J., Dey, B., Berger, E.A., Boyd, M.R., Lifson, J.D., 1999. Cyanovirin–N binds to gp120 to interfere with CD4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does not affect the CD4 binding site on gp120 or soluble CD4- induced conformational changes in gp120. J. Virol. 73, 4360–4371.

Favero, J., 1994. Lectins in AIDS research. Glycobiology 4, 387–396.

Feinberg, H., Mitchell, D.A., Drickamer, K., Weis, W.I., 2001. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. Science 294, 2163–2166.

Fenouillet, E., Clerget-Raslain, B., Gluckman, J.C., Guetard, D., Montagnier, L., Bahraoui, E., 1989. Role of N-linked glycans in the interaction between the envelope glycoprotein of human immunodeficiency virus and its CD4 cellular receptor. Structural enzymatic analysis. J. Exp. Med. 169, 807–822.

Fischer, P.B., Karlsson, G.B., Butters, T.D., Dwek, R.A., Platt, F.M., 1996. *N*-butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry correlates with changes in antibody recognition of the V1/V2 region of gp120. J. Virol. 70, 7143–7152.

Galelli, A., Truffa-Bachi, P., 1993. *Urtica dioica* agglutinin. A superantigenic lectin from stinging nettle rhizome. J. Immunol. 151, 1821–1831.

Gandhi, M.J., Boyd, M.R., Yi, L., Yang, G.G., Vyas, G.N., 2000. Properties of cyanovirin–N (CV–N): inactivation of HIV-1 by sessile cyanovirin–N (sCV–N). Dev. Biol. Stand. 102, 141–148.

Geijtenbeek, T.B., Krooshoop, D.J., Bleijs, D.A., van Vliet, S.J., van Duijnhoven, G.C., Grabovsky, V., Alon, R., Figdor, C.G., van Kooyk, Y., 2000a. DC-SIGN–ICAM-2 interaction mediates dendritic cell trafficking. Nature Immunol. 1, 353–357.

Geijtenbeek, T.B., Kwon, D.S., Torensma, R., van Vliet, S.J., van Duijnhoven, G.C., Middel, J., Cornelissen, I.L., Nottet, H.S., KewalRamani, V.N., Littman, D.R., Figdor, C.G., van Kooyk, Y., 2000b. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. Cell 100, 587–597.

Geijtenbeek, T.B., Torensma, R., van Vliet, S.J., van Duijnhoven, G.C., Adema, G.J., van Kooyk, Y., Figdor, C.G., 2000c. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell 100, 575–585.

Giomarelli, B., Provvedi, R., Meacci, F., Maggi, T., Medaglini, D., Mori, T., McMahon, J.B., Gardella, R., Boyd, M.R., 2002. The microbicide cyanovirin-N expressed on the surface of commensal bacterium *Streptococcus gordonii* captures HIV-1. AIDS 16, 1351–1356.

Goel, M., Jain, D., Kaur, K.J., Kenoth, R., Maiya, B.G., Swamy, M.J., Salunke, D.M., 2001. Functional equality in the absence of structural similarity: an added dimension to molecular mimicry. J. Biol. Chem. 276, 39277–39281.

Gulakowski, R.J., McMahon, J.B., Staley, P.G., Moran, R.A., Boyd, M.R., 1991. A semiautomated multiparameter approach for anti-HIV drug screening. J. Virol. Methods 33, 87–100.

Gustafson, K.R., Sowder, R.C., Henderson, L.E., Cardellina, J.H., McMahon, J.B., Rajamani, U., Pannell, L.K., Boyd, M.R., 1997. Isolation, primary sequence determination, and disulfide bond structure of cyanovirin–N, an anti-HIV (human immunodeficiency virus) protein from the cyanobacterium *Nostoc ellipsosporum*. Biochem. Biophys. Res. Commun. 238, 223–228.

Han, Z., Xiong, C., Mori, T., Boyd, M.R., 2002. Discovery of a stable dimeric mutant of cyanovirin–N (CV–N) from a T7 phage-displayed CV–N mutant library. Biochem. Biophys. Res. Commun. 292, 1036–1043.

Harata, K., Muraki, M., 2000. Crystal structures of *Urtica dioica* agglutinin and its complex with tri-*N*-acetylchitotriose. J. Mol. Biol. 297, 673–681.

Harata, K., Schubert, W.D., Muraki, M., 2001. Structure of *Urtica dioica* agglutinin isolectin I: dimer formation mediated by two zinc ions bound at the sugar-binding site. Acta Crystallogr. D 57, 1513–1517.

Hardman, K.D., Ainsworth, C.F., 1972. Structure of concanavalin A at 2.4 Å resolution. Biochemistry 11, 4910–4919.

Harrop, S.J., Helliwell, J.R., Wan, T.C.M., Kalb (Gilboa), A.J., Tong, L., Yariv, J., 1996. Structure solution of a cubic crystal of concanavalin A complexed with methyl α-*D*-Glucopyranoside. Acta. Crystallogr. D 52, 143–155.

Hart, M.L., Saifuddin, M., Spear, G.T., 2003. Glycosylation inhibitors and neuraminidase enhance human immunodeficiency virus type 1 binding and neutralization by mannose-binding lectin. J. Gen. Virol. 84, 353–360.

Haurum, J.S., Thiel, S., Jones, I.M., Fischer, P.B., Laursen, S.B., Jensenius, J.C., 1993. Complement activation upon binding of mannan-binding protein to HIV envelope glycoproteins. AIDS 7, 1307–1313.

Heilman, C.A., Baltimore, D., 1998. HIV vaccines—where are we going?. Nature Med. 4, 532–534.

Helseth, E., Olshevsky, U., Furman, C., Sodroski, J., 1991. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. J. Virol. 65, 2119–2123.

Hester, G., Wright, C.S., 1996. The mannose-specific bulb lectin from *Galanthus nivalis* (snowdrop) binds mono- and dimannosides at distinct sites. Structure analysis of refined complexes at 2.3 Å and 3.0 Å resolution. J. Mol. Biol. 262, 516–531.

Hester, G., Kaku, H., Goldstein, I.J., Wright, C.S., 1995. Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. Nature Struct. Biol. 2, 472–479.

Ho, D.D., McKeating, J.A., Li, X.L., Moudgil, T., Daar, E.S., Sun, N.C., Robinson, J.E., 1991. Conformational epitope on gp120 important in CD4 binding and human immunodeficiency virus type 1 neutralization identified by a human monoclonal antibody. J. Virol. 65, 489–493.

Igarashi, T., Brown, C., Azadegan, A., Haigwood, N., Dimitrov, D., Martin, M.A., Shibata, R., 1999. Human immunodeficiency virus type 1 neutralizing antibodies accelerate clearance of cell-free virions from blood plasma. Nature Med. 5, 211–216.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T., Yamashina, I., 1987. Serum lectin with known structure activates complement through the classical pathway. J. Biol. Chem. 262, 7451–7454.

Iobst, S.T., Wormald, M.R., Weis, W.I., Dwek, R.A., Drickamer, K., 1994. Binding of sugar ligands to Ca(2+)-dependent animal lectins. I. Analysis of mannose binding by site-directed mutagenesis and NMR. J. Biol. Chem. 269, 15505–15511.

Jack, D.L., Klein, N.J., Turner, M.W., 2001. Mannose-binding lectin: targeting the microbial world for complement attack and opsonophagocytosis. Immunol. Rev. 180, 86–99.

Jeyaprakash, A.A., Geetha, R.P., Banuprakash, R.G., Banumathi, S., Betzel, C., Sekar, K., Surolia, A., Vijayan, M., 2002. Crystal structure of the jacalin–T-antigen complex and a comparative study of lectin–T-antigen complexes. J. Mol. Biol. 321, 637–645.

Jeyaprakash, A.A., Katiyar, S., Swaminathan, C.P., Sekar, K., Surolia, A., Vijayan, M., 2003. Structural basis of the carbohydrate specificities of jacalin: an X-ray and modeling study. J. Mol. Biol. 332, 217–228.

Kaku, H., Goldstein, I.J., 1992. Interaction of linear manno-oligosaccharides with three mannose-specific bulb lectins. Comparison with mannose/glucose-binding lectins. Carbohydr. Res. 229, 337–346.

Kaku, H., Van Damme, E.J., Peumans, W.J., Goldstein, I.J., 1990. Carbohydrate-binding specificity of the daffodil (*Narcissus pseudonarcissus*) and amaryllis (*Hippeastrum hybr.*) bulb lectins. Arch. Biochem. Biophys. 279, 298–304.

Kaku, H., Goldstein, I.J., Oscarson, S., 1991. Interactions of five *D*-mannose-specific lectins with a series of synthetic branched trisaccharides. Carbohydr. Res. 213, 109–116.

Kalb, A.J., Levitzki, A., 1968. Metal-binding sites of concanavalin A and their role in the binding of alpha-methyl *D*-glucopyranoside. Biochem. J. 109, 669–672.

Kantardjieff, K.A., Hochtl, P., Segelke, B.W., Tao, F.M., Rupp, B., 2002. Concanavalin A in a dimeric crystal form: revisiting structural accuracy and molecular flexibility. Acta Crystallogr. D 58, 735–743.

Karlsson, G.B., Butters, T.D., Dwek, R.A., Platt, F.M., 1993. Effects of the imino sugar *N*-butyldeoxynojirimycin on the *N*-glycosylation of recombinant gp120. J. Biol. Chem. 268, 570–576.

Kelley, B.S., Chang, L.C., Bewley, C.A., 2002. Engineering an obligate domain-swapped dimer of cyanovirin–N with enhanced anti-HIV activity. J. Am. Chem. Soc. 124, 3210–3211.

Kilby, J.M., Hopkins, S., Venetta, T.M., DiMassimo, B., Cloud, G.A., Lee, J.Y., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., Matthews, T., Johnson, M.R., Nowak, M.A., Shaw, G.M., Saag, M.S., 1998. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. Nature Med. 4, 1302–1307.

Kornfeld, R., Kornfeld, S., 1985. Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. 54, 631–664.

Kunert, R., Ruker, F., Katinger, H., 1998. Molecular characterization of five neutralizing anti-HIV type 1 antibodies: identification of nonconventional D segments in the human monoclonal antibodies 2G12 and 2F5. AIDS Res. Hum. Retroviruses 14, 1115–1128.

Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A., 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393, 648–659.

Kwong, P.D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R.W., Sodroski, J., Hendrickson, W.A., 2000. Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. Struct. Fold. Des. 8, 1329–1339.

Lange, J.M., Karam, M., Piot, P., 1993. Boost for vaginal microbicides against HIV. Lancet 342, 1356.

Lee, R.T., Ichikawa, Y., Kawasaki, T., Drickamer, K., Lee, Y.C., 1992. Multivalent ligand binding by serum mannose-binding protein. Arch. Biochem. Biophys. 299, 129–136.

Lee, S.J., Evers, S., Roeder, D., Parlow, A.F., Risteli, J., Risteli, L., Lee, Y.C., Feizi, T., Langen, H., Nussenzweig, M.C., 2002. Mannose receptor-mediated regulation of serum glycoprotein homeostasis. Science 295, 1898–1901.

Lee, Y.C., Lee, R.T., Rice, K., Ichikawa, Y., Wong, T.-C., 1991. Topography of binding sites of animal lectins: ligands' view. Pure Appl. Chem. 63, 499–506.

Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N., Gregory, T.J., 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J. Biol. Chem. 265, 10373–10382.

Li, H., Wang, L.X., 2004. Design and synthesis of a template-assembled oligomannose cluster as an epitope mimic for human HIV-neutralizing antibody 2G12. Org. Biomol. Chem. 2, 483–488.

Lifson, J., Coutre, S., Huang, E., Engleman, E., 1986. Role of envelope glycoprotein carbohydrate in human immunodeficiency virus (HIV) infectivity and virus-induced cell fusion. J. Exp. Med. 164, 2101–2106.

Liu, Y., Eisenberg, D., 2002. 3D domain swapping: as domains continue to swap. Protein Sci. 11, 1285–1299.

Lu, M., Blacklow, S.C., Kim, P.S., 1995. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. Nature Struct. Biol. 2, 1075–1082.

MacArthur, M.W., Thornton, J.M., 1991. Influence of proline residues on protein conformation. J. Mol. Biol. 218, 397–412.

Mahanta, S.K., Sanker, S., Rao, N.V., Swamy, M.J., Surolia, A., 1992. Primary structure of a Thomsen–Friedenreich-antigen-specific lectin, jacalin [*Artocarpus integrifolia* (jack fruit) agglutinin]. Evidence for the presence of an internal repeat. Biochem. J. 284 (Part 1), 95–101.

Mandal, D.K., Kishore, N., Brewer, C.F., 1994. Thermodynamics of lectin–carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. Biochemistry 33, 1149–1156.

Mariner, J.M., McMahon, J.B., O'Keefe, B.R., Nagashima, K., Boyd, M.R., 1998. The HIV-inactivating protein, Cyanovirin-N, does not block gp120- mediated virus-to-cell binding. Biochem. Biophys. Res. Commun. 248, 841–845.

Mascola, J.R., Stiegler, G., VanCott, T.C., Katinger, H., Carpenter, C.B., Hanson, C.E., Beary, H., Hayes, D., Frankel, S.S., Birx, D.L., Lewis, M.G., 2000. Protection of macaques against vaginal transmission of a pathogenic HIV-1/SIV chimeric virus by passive infusion of neutralizing antibodies. Nature Med. 6, 207–210.

Matthews, T.J., Weinhold, K.J., Lyerly, H.K., Langlois, A.J., Wigzell, H., Bolognesi, D.P., 1987. Interaction between the human T-cell lymphotropic virus type IIIB envelope glycoprotein gp120 and the surface antigen CD4: role of carbohydrate in binding and cell fusion. Proc. Natl. Acad. Sci. USA 84, 5424–5428.

Meritt, E.A., Murphy, M.E.P., 1994. *Raster3D* Version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr. D 50, 869–873.

Mitchell, D.A., Fadden, A.J., Drickamer, K., 2001. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. J. Biol. Chem. 276, 28939–28945.

Mizuno, Y., Kozutsumi, Y., Kawasaki, T., Yamashina, I., 1981. Isolation and characterization of a mannan-binding protein from rat liver. J. Biol. Chem. 256, 4247–4252.

Mizuochi, T., Spellman, M.W., Larkin, M., Solomon, J., Basa, L.J., Feizi, T., 1988. Carbohydrate structures of the human immunodeficiency virus (HIV) recombinant envelope glycoprotein gp120 produced in Chinese-hamster ovary cells. Biochem. J. 254, 599–603.

Mizuochi, T., Spellman, M.W., Larkin, M., Solomon, J., Basa, L.J., Feizi, T., 1988. Structural characterization by chromatographic profiling of the oligosaccharides of human immunodeficiency virus (HIV) recombinant envelope glycoprotein gp120 produced in Chinese hamster ovary cells. Biomed. Chromatogr. 2, 260–270.

Modrow, S., Hahn, B.H., Shaw, G.M., Gallo, R.C., Wong-Staal, F., Wolf, H., 1987. Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J. Virol. 61, 570–578.

Moore, J.P., Sodroski, J., 1996. Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein. J. Virol. 70, 1863–1872.

Moothoo, D.N., Naismith, J.H., 1998. Concanavalin A distorts the beta-GlcNAc-(1→2)-Man linkage of beta-GlcNAc-(1→2)-alpha-Man-(1→3)-[beta-GlcNAc-(1→2)-alpha-Man-(1→6)]-Man upon binding. Glycobiology 8, 173–181.

Mori, T., Boyd, M.R., 2001. Cyanovirin–N, a potent human immunodeficiency virus-inactivating protein, blocks both CD4-dependent and CD4-independent binding of soluble gp120 (sgp120) to target cells, inhibits sCD4-induced binding of sgp120 to cell-associated CXCR4, and dissociates bound sgp120 from target cells. Antimicrob. Agents Chemother. 45, 664–672.

Mori, T., Shoemaker, R.H., Gulakowski, R.J., Krepps, B.L., McMahon, J.B., Gustafson, K.R., Pannell, L.K., Boyd, M.R., 1997a. Analysis of sequence requirements for biological activity of cyanovirin–N, a potent HIV (human immunodeficiency virus)-inactivating protein. Biochem. Biophys. Res. Commun. 238, 218–222.

Mori, T., Shoemaker, R.H., McMahon, J.B., Gulakowski, R.J., Gustafson, K.R., Boyd, M.R., 1997b. Construction and enhanced cytotoxicity of a [Cyanovirin–N]-[*Pseudomonas* exotoxin] conjugate against human immunodeficiency virus-infected cells. Biochem. Biophys. Res. Commun. 239, 884–888.

Mori, T., Gustafson, K.R., Pannell, L.K., Shoemaker, R.H., Wu, L., McMahon, J.B., Boyd, M.R., 1998. Recombinant production of Cyanovirin-N, a potent human immunodeficiency virus-inactivating protein derived from a cultured cyanobacterium. Protein Exp. Purif. 12, 151–158.

Mullin, N.P., Hall, K.T., Taylor, M.E., 1994. Characterization of ligand binding to a carbohydrate-recognition domain of the macrophage mannose receptor. J. Biol. Chem. 269, 28405–28413.

Naismith, J.H., Field, R.A., 1996. Structural basis of trimannoside recognition by concanavalin A. J. Biol. Chem. 271, 972–976.

Naismith, J.H., Emmerich, C., Habash, J., Harrop, S.J., Helliwell, J.R., Hunter, W.N., Raftery, J., Kalb, A.J., Yariv, J., 1994. Refined structure of concanavalin a complexed with methyl α-*D*-Mannopyranoside at 2.0 Å resolution and comparison with the saccharide-free structure. Acta Crystallogr. D 50, 847–858.

Ng, K.K., Drickamer, K., Weis, W.I., 1996. Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. J. Biol. Chem. 271, 663–674.

Ng, K.K., Kolatkar, A.R., Park-Snyder, S., Feinberg, H., Clark, D.A., Drickamer, K., Weis, W.I., 2002. Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. J. Biol. Chem. 277, 16088–16095.

Nicholls, A., 1992. GRASP: Graphical Representation and Analysis of Surface Properties. Columbia University, New York, NY.

Nixon, D.F., Broliden, K., Ogg, G., Broliden, P.A., 1992. Cellular and humoral antigenic epitopes in HIV and SIV. Immunology 76, 515–534.

Nussbaum, O., Broder, C.C., Berger, E.A., 1994. Fusogenic mechanisms of enveloped-virus glycoproteins analyzed by a novel recombinant vaccinia virus-based assay quantitating cell fusion-dependent reporter gene activation. J. Virol. 68, 5411–5422.

O'Keefe, B.R., 2001. Biologically active proteins from natural product extracts. J. Natl. Prod. 64, 1373–1381.

O'Keefe, B.R., Shenoy, S.R., Xie, D., Zhang, W., Muschik, J.M., Currens, M.J., Chaiken, I., Boyd, M.R., 2000. Analysis of the interaction between the HIV-inactivating protein Cyanovirin–N and soluble forms of the envelope glycoproteins gp120 and gp41. Mol. Pharmacol. 58, 982–992.

Petersen, S.V., Thiel, S., Jensenius, J.C., 2001. The mannan-binding lectin pathway of complement activation: biology and disease association. Mol. Immunol. 38, 133–149.

Peterson, A., Seed, B., 1988. Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4. Cell 54, 65–72.

Petrescu, A.J., Petrescu, S.M., Dwek, R.A., Wormald, M.R., 1999. A statistical analysis of N- and O-glycan linkage conformations from crystallographic data. Glycobiology 9, 343–352.

Peumans, W.J., De Ley, M., Broekaert, W.F., 1984. An unusual lectin from stinging nettle (*Urtica dioica*) rhizomes. FEBS Lett. 177, 99–103.

Pinter, A., Honnen, W.J., Tilley, S.A., Bona, C., Zaghouani, H., Gorny, M.K., Zolla-Pazner, S., 1989. Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1. J. Virol. 63, 2674–2679.

Pohlmann, S., Baribaud, F., Doms, R.W., 2001a. DC-SIGN and DC-SIGNR: helping hands for HIV. Trends Immunol. 22, 643–646.

Pohlmann, S., Baribaud, F., Lee, B., Leslie, G.J., Sanchez, M.D., Hiebenthal-Millow, K., Munch, J., Kirchhoff, F., Doms, R.W., 2001b. DC-SIGN interactions with human immunodeficiency virus type 1 and 2 and simian immunodeficiency virus. J. Virol. 75, 4664–4672.

Pohlmann, S., Soilleux, E.J., Baribaud, F., Leslie, G.J., Morris, L.S., Trowsdale, J., Lee, B., Coleman, N., Doms, R.W., 2001c. DC-SIGNR, a DC-SIGN homologue expressed in endothelial cells, binds to human and simian immunodeficiency viruses and activates infection in trans. Proc. Natl. Acad. Sci. USA 98, 2670–2675.

Posner, M.R., Hideshima, T., Cannon, T., Mukherjee, M., Mayer, K.H., Byrn, R.A., 1991. An IgG human monoclonal antibody that reacts with HIV–1/GP120, inhibits virus binding to cells, and neutralizes infection. J. Immunol. 146, 4325–4332.

Rao, V.S.R., Qasba, P.K., Balaji, P.V., Chandrasekaran, R., 1998. Conformation of Carbohydrates. Harwood Academic Publishers, The Netherlands.

Reeke Jr., G.N., Becker, J.W., Edelman, G.M., 1975. The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. J. Biol. Chem. 250, 1525–1547.

Reeke Jr., G.N., Becker, J.W., Edelman, G.M., 1978. Changes in the three-dimensional structure of concanavalin A upon demetallization. Proc. Natl. Acad. Sci. USA 75, 2286–2290.

Rice, K.G., Lee, Y.C., 1993. Oligosaccharide valency and conformation in determining binding to the asialoglycoprotein receptor of rat hepatocytes. Adv. Enzymol. Relat. Areas Mol. Biol. 66, 41–83.

Rousseau, F., Schymkowitz, J.W., Wilkinson, H.R., Itzhaki, L.S., 2001. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. Proc. Natl. Acad. Sci. USA 98, 5596–5601.

Rovira, P., Buckle, M., Abastado, J.P., Peumans, W.J., Truffa-Bachi, P., 1999. Major histocompatibility class I molecules present *Urtica dioica* agglutinin, a superantigen of vegetal origin, to T lymphocytes. Eur. J. Immunol. 29, 1571–1580.

Rudd, P.M., Dwek, R.A., 1997. Glycosylation: heterogeneity and the 3D structure of proteins. Crit. Rev. Biochem. Mol. Biol. 32, 1–100.

Saifuddin, M., Hart, M.L., Gewurz, H., Zhang, Y., Spear, G.T., 2000. Interaction of mannose-binding lectin with primary isolates of human immunodeficiency virus type 1. J. Gen. Virol. 81, 949–955.

Sanders, R.W., Venturi, M., Schiffner, L., Kalyanaraman, R., Katinger, H., Lloyd, K.O., Kwong, P.D., Moore, J.P., 2002. The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. J. Virol. 76, 7293–7305.

Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A., Vijayan, M., 1996. A novel mode of carbohydrate recognition in jacalin, a *Moraceae* plant lectin with a beta-prism fold. Nature Struct. Biol. 3, 596–603.

Sansom, C., Wlodawer, A., 2000. Drugs targeted at HIV—successes and resistance. In: Rodrigo, A., Learn, G.H. (Eds.), Computational Analysis of Human Immunodeficiency Virus Molecular Sequences. Kluwer Academic Publishers, Boston, MA, pp. 269–286.

Sastry, M.V., Banarjee, P., Patanjali, S.R., Swamy, M.J., Swarnalatha, G.V., Surolia, A., 1986. Analysis of saccharide binding to *Artocarpus integrifolia* lectin reveals specific recognition of T-antigen (beta-D-Gal(1→3)D-GalNAc). J. Biol. Chem. 261, 11726–11733.

Sauerborn, M.K., Wright, L.M., Reynolds, C.D., Grossmann, J.G., Rizkallah, P.J., 1999. Insights into carbohydrate recognition by *Narcissus pseudonarcissus* lectin: the crystal structure at 2 Å resolution in complex with alpha1-3 mannobiose. J. Mol. Biol. 290, 185–199.

Saul, F.A., Rovira, P., Boulot, G., Damme, E.J., Peumans, W.J., Truffa-Bachi, P., Bentley, G.A., 2000. Crystal structure of *Urtica dioica* agglutinin, a superantigen presented by MHC molecules of class I and class II. Struct. Fold. Des. 8, 593–603.

Scanlan, C.N., Pantophlet, R., Wormald, M.R., Ollmann, S.E., Stanfield, R., Wilson, I.A., Katinger, H., Dwek, R.A., Rudd, P.M., Burton, D.R., 2002. The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1→2 mannose residues on the outer face of gp120. J. Virol. 76, 7306–7321.

Schulke, N., Vesanen, M.S., Sanders, R.W., Zhu, P., Lu, M., Anselma, D.J., Villa, A.R., Parren, P.W., Binley, J.M., Roux, K.H., Maddon, P.J., Moore, J.P., Olson, W.C., 2002. Oligomeric and conformational properties of a proteolytically mature, disulfide-stabilized human immunodeficiency virus type 1 gp140 envelope glycoprotein. J. Virol. 76, 7760–7776.

Sharon, N., 1993. Lectin–carbohydrate complexes of plants and animals: an atomic view. Trends Biochem. Sci. 18, 221–226.

Sharon, N., Lis, H., 1991. Legume lectins—a large family of homologous proteins. FASEB J. 4, 3198–3208.

Shenoy, S.R., O'Keefe, B.R., Bolmstedt, A.J., Cartner, L.K., Boyd, M.R., 2001. Selective interactions of the human immunodeficiency virus-inactivating protein Cyanovirin–N with high-mannose oligosaccharides on gp120 and other glycoproteins. J. Pharmacol. Exp. Ther. 297, 704–710.

Shibuya, N., Goldstein, I.J., Shafer, J.A., Peumans, W.J., Broekaert, W.F., 1986. Carbohydrate binding properties of the stinging nettle (*Urtica dioica*) rhizome lectin. Arch. Biochem. Biophys. 249, 215–224.

Shibuya, N., Goldstein, I.J., Van Damme, E.J., Peumans, W.J., 1988. Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. J. Biol. Chem. 263, 728–734.

Shoham, M., Yonath, A., Sussman, J.L., Moult, J., Traub, W., Kalb, A.J., 1979. Crystal structure of demetallized concanavalin A: the metal-binding region. J. Mol. Biol. 131, 137–155.

Shugars, D.C., Wild, C.T., Greenwell, T.K., Matthews, T.J., 1996. Biophysical characterization of recombinant proteins expressing the leucine zipper-like domain of the human immunodeficiency virus type 1 transmembrane protein gp41. J. Virol. 70, 2982–2991.

Thali, M., Furman, C., Ho, D.D., Robinson, J., Tilley, S., Pinter, A., Sodroski, J., 1992. Discontinuous, conserved neutralization epitopes overlapping the CD4-binding region of human immunodeficiency virus type 1 gp120 envelope glycoprotein. J. Virol. 66, 5635–5641.

Thali, M., Moore, J.P., Furman, C., Charles, M., Ho, D.D., Robinson, J., Sodroski, J., 1993. Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120–CD4 binding. J. Virol. 67, 3978–3988.

Thiel, S., Vorup-Jensen, T., Stover, C.M., Schwaeble, W., Laursen, S.B., Poulsen, K., Willis, A.C., Eggleton, P., Hansen, S., Holmskov, U., Reid, K.B., Jensenius, J.C., 1997. A second serine protease associated with mannan-binding lectin that activates complement. Nature 386, 506–510.

Trkola, A., Pomales, A.B., Yuan, H., Korber, B., Maddon, P.J., Allaway, G.P., Katinger, H., Barbas III, C.F., Burton, D.R., Ho, D.D., 1995. Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4–IgG. J. Virol. 69, 6609–6617.

Trkola, A., Dragic, T., Arthos, J., Binley, J.M., Olson, W.C., Allaway, G.P., Cheng-Mayer, C., Robinson, J., Maddon, P.J., Moore, J.P., 1996a. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. Nature 384, 184–187.

Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J.P., Katinger, H., 1996b. Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. J. Virol. 70, 1100–1108.

UNAIDS-WHO. AIDS epidemic update: 2003. http://www.unaids.org.

Van Damme, E.J.M., Peumanns, W.J., 1990. Isolectins in *Narcissus*: complexity, inter- and intraspecies differences and development control. Physiol. Plant. 79, 1–6.

Van Damme, E.J.M., Allen, A.K., Peumans, W.J., 1987. Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. FEBS Lett. 215, 140–144.

Van Damme, E.J.M., Allen, A.K., Peumans, W.J., 1988a. Related mannose-specific lectins from different species of the family *Amaryllidaceae*. Physiol. Plant. 73, 52–57.

Van Damme, E.J.M., Broekaert, W.F., Peumans, W.J., 1988b. The *Urtica dioica* agglutinin is a complex mixture of isolectins. Plant Physiol. 86, 598–601.

Van Damme, E.J.M., Goldstein, I.J., Peumanns, W.J., 1991. A comparative study of mannose-binding lectins from the *Amaryllidaceae* and *Alliaceae*. Phytochemisty 30, 509–514.

Van Damme, E.J.M., Goldstein, I.J., Vercammen, G.V.J., Peumanns, W.J., 1992. Lectins of members of the *Amaryllidaceae* are encoded by multigene families which show extensive homology. Physiol. Plant 86, 245–252.

van Emmerik, L.C., Kuijper, E.J., Fijen, C.A., Dankert, J., Thiel, S., 1994. Binding of mannan-binding protein to various bacterial pathogens of meningitis. Clin. Exp. Immunol. 97, 411–416.

Weiler, B.E., Schroder, H.C., Stefanovich, V., Stewart, D., Forrest, J.M., Allen, L.B., Bowden, B.J., Kreuter, M.H., Voth, R., Muller, W.E., 1990. Sulphoevernan, a polyanionic polysaccharide, and the narcissus lectin potently inhibit human immunodeficiency virus infection by binding to viral envelope protein. J. Gen. Virol. 71 (Part 9), 1957–1963.

Weis, W.I., Drickamer, K., 1994. Trimeric structure of a C-type mannose-binding protein. Structure 2, 1227–1240.

Weis, W.I., Drickamer, K., 1996. Structural basis of lectin–carbohydrate recognition. Annu. Rev. Biochem. 65, 441–473.

Weis, W.I., Kahn, R., Fourme, R., Drickamer, K., Hendrickson, W.A., 1991. Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. Science 254, 1608–1615.

Weis, W.I., Drickamer, K., Hendrickson, W.A., 1992. Structure of a C-type mannose-binding protein complexed with an oligosaccharide. Nature 360, 127–134.

Weis, W.I., Taylor, M.E., Drickamer, K., 1998. The C-type lectin superfamily in the immune system. Immunol. Rev. 163, 19–34.

Weiss, C.D., Levy, J.A., White, J.M., 1990. Oligomeric organization of gp120 on infectious human immunodeficiency virus type 1 particles. J. Virol. 64, 5674–5677.

Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C., 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature 387, 426–430.

Willey, R.L., Rutledge, R.A., Dias, S., Folks, T., Theodore, T., Buckler, C.E., Martin, M.A., 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. Proc. Natl. Acad. Sci. USA 83, 5038–5042.

Wilson, I.A., Skehel, J.J., Wiley, D.C., 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289, 366–373.

Wood, S.D., Wright, L.M., Reynolds, C.D., Rizkallah, P.J., Allen, A.K., Peumans, W.J., Van Damme, E.J., 1999. Structure of the native (unligated) mannose-specific bulb lectin from *Scilla campanulata* (bluebell) at 1.7 Å resolution. Acta Crystallogr. D 55 (Part 7), 1264–1272.

Woods, R.J., Pathiaseril, A., Wormald, M.R., Edge, C.J., Dwek, R.A., 1998. The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. Eur. J. Biochem. 258, 372–386.

Wright, C.S., 1977. The crystal structure of wheat germ agglutinin at 2-2 Å resolution. J. Mol. Biol. 111, 439–457.

Wright, C.S., Hester, G., 1996. The 2.0 Å structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding modes. Structure 4, 1339–1352.

Wright, L.M., Wood, S.D., Reynolds, C.D., Rizkallah, P.J., Allen, A.K., Peumans, W.J., 1997. X-ray crystallographic studies of two native monocot lectins, and three carbohydrate-bound complexes. Eur. J. Cell Biol. 74, 114.

Wyatt, R., Sodroski, J., 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science 280, 1884–1888.

Wyatt, R., Kwong, P.D., Desjardins, E., Sweet, R.W., Robinson, J., Hendrickson, W.A., Sodroski, J.G., 1998. The antigenic structure of the HIV gp120 envelope glycoprotein. Nature 393, 705–711.

Yang, H., Czapla, T.H., 1993. Isolation and characterization of cDNA clones encoding jacalin isolectins. J. Biol. Chem. 268, 5905–5910.

Yang, F., Bewley, C.A., Louis, J.M., Gustafson, K.R., Boyd, M.R., Gronenborn, A.M., Clore, G.M., Wlodawer, A., 1999. Crystal structure of Cyanovirin–N, a potent HIV-inactivating protein, shows unexpected domain swapping. J. Mol. Biol. 288, 403–412.

Yariv, J., Kalb, A.J., Levitzki, A., 1968. The interaction of concanavalin A with methyl alpha-*D*-glucopyranoside. Biochim. Biophys. Acta 165, 303–305.

Zhou, G., Ferrer, M., Chopra, R., Kapoor, T.M., Strassmaier, T., Weissenhorn, W., Skehel, J.J., Oprian, D., Schreiber, S.L., Harrison, S.C., Wiley, D.C., 2000. The structure of an HIV-1 specific cell entry inhibitor in complex with the HIV-1 gp41 trimeric core. Bioorg. Med. Chem. 8, 2219–2227.

Zhu, X., Borchers, C., Bienstock, R.J., Tomer, K.B., 2000. Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. Biochemistry 39, 11194–11204.